(19) Europäisches Patentamt
European Patent Office
Office européen des br vets

(11) EP 0 849 359 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
24.06.1998 Bulletin 1998/26

(51) Int. Cl.[6]: **C12N 9/10**, C12N 15/54

(21) Application number: **97122417.5**

(22) Date of filing: **18.12.1997**

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **18.12.1996 JP 338673/96**

(71) Applicant:
Sumitomo Chemical Company, Limited
Osaka-shi, Osaka-fu (JP)

(72) Inventors:
- **Wantanabe, Eijiro**
  Takarazuka-shi, Hyogo-ken (JP)
- **Oeda, Kenji**
  Sakyo-ku, Kyoto-shi (JP)

(74) Representative:
VOSSIUS & PARTNER
Siebertstrasse 4
81675 München (DE)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Raffinose synthetase genes and the use thereof

(57) Raffinose synthetase genes coding for proteins capable of producing raffinose by combining a D-galactosyl group through an α(1→6) bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule were isolated from various plants. These raffinose synthetase genes are useful to change the content of raffinose family oligosaccharides in plants.

EP 0 849 359 A2

Description

FIELD OF THE INVENTION

The present invention relates to raffinose synthetase genes and their use.

BACKGROUND OF THE INVENTION

Raffinose family oligosaccharides are derivatives of sucrose, which are represented by o-α-D-galactopyranosyl-(1→6) n-o-α-D-glucopyranosyl-(1→2)-β-D-fluctofuranoside as the general formula, and they are designated "raffinose" when n = 1, "stachyose" when n = 2, "verbascose" when n = 3, and "ajugose" when n = 4.

The greatest contents of such raffinose family oligosaccharides are found in plants, except for sucrose, and it has been known that they are contained not only in higher plants including gymnosperms such as pinaceous plants (e.g., spruce) and angiosperms such as leguminous plants (e.g., soybean, kidney bean), brassicaceous plants (e.g., rape), chenopodiaceous plants (e.g., sugar beet), malvaceous plants (e.g., cotton), and salicaceous plants (e.g., poplar), but also in green algae, chlorella. Thus, they occur widely in the plant kingdom similarly to sucrose.

Raffinose family oligosaccharides play a role as reserve sugars in the storage organs or seeds of many plants or as translocating sugars in the phenomenon of sugar transportation between the tissues of some plants.

Furthermore, it has been known that raffinose family oligosaccharides have an effect of giving good conditions of nterobacterial flora, if present at a suitable amount in food. Therefore, raffinose family oligosaccharides have already been used as a functional food material for addition to some kinds of food and utilized in the field of specified healthful food.

Raffinose family oligosaccharides having such a role and utility are produced by the raffinose oligosaccharide synthesis system beginning with sucrose in many plants. This biosynthesis system usually involves a reaction for the sequential addition of galactosyl groups from galactotinol through an α(1→6) bond to a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

In the first step of this biosynthesis system, raffinose synthetase is an enzyme concerned in the reaction of raffinose production by combining a D-galactosyl group from galactotinol through an α(1→6) bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule. It has been suggested that this enzyme constitutes a rate limiting step in the above synthesis system, and it has been revealed that this enzyme is quite important in the control of biosynthesis of raffinose family oligosaccharides.

The control of expression level or activity of raffinose synthetase in plants makes it possible to change the contents of raffinose family oligosaccharides in these plants. However, raffinose synthetase, although the presence of this nzyme itself was already confirmed in many plants by the measurement of its activity with a biochemical technique, has not yet been successfully isolated and purified as a homogeneous protein. In addition, the amino acid sequence of this enzyme is still unknown, and no report has been made on an attempt at beginning to isolate a gene coding for this enzyme.

SUMMARY OF THE INVENTION

Under these circumstances, the present inventors have intensively studied and finally succeeded in isolating a raffinose synthetase and a gene coding for this enzyme from broad bean, thereby completing the present invention.

Thus, the present invention provides the following:

1) A raffinose synthetase gene isolated from a plant and having a nucleotide sequence coding for an amino acid sequence of a protein capable of producing raffinose by combining a D-galactosyl group through an α(1→6) bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

2) The raffinose synthetase gene according to item 1, wherein the plant is a dicotyledon.

3) The raffinose synthetase gene according to item 2, wherein the dicotyledon is a leguminous plant.

4) The raffinose synthetase gene according to item 3, wherein the leguminous plant is broad bean.

5) A raffinose synthetase gene having a nucleotide sequence coding for protein (a) or (b) as defined below:

(a) protein having the amino acid sequence of SEQ ID NO:1;

(b) protein having an amino acid sequence derived by deletion, replacement, modification or addition of one or several amino acids in the amino acid sequence of SEQ ID NO:1, and capable of producing raffinose by combining a D-galactosyl group through an α(1→6) bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

6) A raffinose synthetase gene having the nucleotide sequence f SEQ ID NO:2.

7) The raffinose synthetase gene according to item 3, wherein the leguminous plant is soybean.

8) A raffinos synthetase gen having a nucleotide sequence coding for protein (a) or (b) as defined below:

(a) protein having the amino acid sequence of SEQ ID NO:3;

(b) protein having an amino acid sequence derived by deletion, replacement, modification or addition of one or several amino acids in the amino acid sequence of SEQ ID NO:3, and capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

9) A raffinose synthetase gene having the nucleotide sequence of SEQ ID NO:4.

10) The raffinose synthetase gene according to item 2, wherein the dicotyledon is a lamiaceous plant.

11) The raffinose synthetase gene according to item 10, wherein the lamiaceous plant is Japanese artichoke.

12) A raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of SEQ ID NO:5.

13) A raffinose synthetase gene having the nucleotide sequence of SEQ ID NO:6.

14) The raffinose synthetase gene according to item 1, wherein the plant is a monocotyledon.

15) The raffinose synthetase gene according to item 14, wherein the mono-cotyledon is a gramineous plant.

16) The raffinose synthetase gene according to item 15, wherein the gramineous plant is corn.

17) A raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of SEQ ID NO:7.

18) A raffinose synthetase gene having the nucleotide sequence of SEQ ID NO:8.

19) A raffinose synthetase protein having amino acid sequence (a) or (b) as defined below:

(a) amino acid sequence of SEQ ID NO:1 or SEQ ID NO:3;

(b) amino acid sequence derived by deletion, replacement, modification or addition of one or several amino acids in the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:3;

the protein being capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

20) A raffinose synthetase protein having the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

21) A gene fragment having a partial nucleotide sequence of the raffinose synthetase gene of item 1, 2, 3, 4, 7, 10, 11, 14, 15 or 16.

22) A gene fragment having a partial nucleotide sequence of the raffinose synthetase gene of item 5, 6, 8, 9, 12, 13, 17 or 18.

23) The gene fragment according to item 21 or 22, wherein the number of nucleotides is in the range of from 15 to 50.

24) A method for the detection of a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof, which comprises hybridizing a probe of the labeled gene fragment of item 21, 22 or 23 to an organism-derived genomic DNA or cDNA fragment; and detecting the DNA fragment bound specifically to the probe.

25) A method for the detection of a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof, which comprises hybridizing a probe of the labeled gene fragment of item 21, 22 or 23 to a plant-derived genomic DNA or cDNA fragment; and detecting the DNA fragment bound specifically to the probe.

26) A method for the amplification of a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof, which comprises annealing a primer having a nucleotide sequence of the gene fragment of item 21, 22 or 23 to organism-derived genomic DNA or cDNA; and amplifying the resulting DNA fragment by polymerase chain reaction.

27) A method for the amplification of a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof, which comprises annealing a primer having a nucleotide sequence of the gene fragment of item 21, 22 or 23 to plant-derived genomic DNA or cDNA; and amplifying the resulting DNA fragment by polymerase chain reaction.

28) A method for obtaining a raffinose synthetase gene, comprising the steps of identifying a DNA fragment containing a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof by the method of item 24, 25, 26 or 27; and isolating and purifying the DNA fragment identified.

29) A raffinose synthetase gene obtained by identifying a DNA fragm nt containing a raffinose synthetase gen or a gen fragment having a partial nucleotide sequence thereof by the method of item 24, 25, 26 or 27; and isolating

and purifying the DNA fragment id ntified.

30) A chimera gene comprising the raffinose synthetase gene of item 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or 29, and a promoter linked thereto.

31) A transformant obtained by introducing the chimera gene of item 30 into a host organism.

32) A plasmid comprising the raffinose synthetase gene of item 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29 or 30.

33) A host organism transformed with the plasmid of item 32, or a cell thereof.

34) A microorganism transformed with the plasmid of item 32.

35) A plant transformed with the plasmid of item 32, or a cell thereof.

36) A method for metabolic modification, which comprises introducing the raffinose synthetase gene of item 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29 or 30 into a host organism or a cell thereof, so that the content of raffinose family oligosaccharides in the host organism or the cell thereof is changed.

37) A method for the production of a raffinose synthetase protein, which comprises isolating and purifying a raffinose synthetase protein from a culture obtained by cultivating the microorganism of item 34.

38) An anti-raffinose synthetase antibody capable of binding to the raffinose synthetase protein of item 19 or 20.

39) A method for the detection of a raffinose synthetase protein, which comprises treating a test protein with the anti-raffinose synthetase antibody of item 38; and detecting the raffinose synthetase protein by antigen-antibody reaction between the antibody and the raffinose synthetase protein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the construction of plasmids used for the expression of a raffinose synthetase gene in *Escherichia coli*. pBluescriptKS-RS is a plasmid containing the raffinose synthetase gene cloned therein. RS represents the raffinose synthetase gene, and the nucleotide sequences shown in the upper portion of this figure are those of both terminal portions of the raffinose synthetase gene. A partial sequence represented by small letters is a nucleotide sequence derived from the vector pBluescriptII KS-. Two boxed nucleotide sequences are the initiation codon (ATG) and termination codon (TGATAA) of the raffinose synthetase gene, respectively. The recognition sites for several restriction endonucleases are shown above the nucleotide sequences. pGEX-RS and pTrc-RS are plasmids used for the expression of the raffinose synthetase gene in *E. coli*. Ptac, Ptrc, GST, $lacI^q$, and rrnB represent tac promoter, trc promoter, glutathione-S-transferase gene, lactose repressor gene, and termination signal for the transcription of ribosomal RNA, respectively.

Figure 2 shows the construction of expression vectors used for the expression in plants of chimera genes each having a raffinose synthetase gene and a promoter linked thereto. The restriction endonuclease map of the raffinose synthetase gene cloned in the plasmid pBluescriptKS-RS is shown in the lower portion of this figure. pBI221RS and pBI221(-)RS indicate the restriction endonuclease maps of expression vectors used for the transformation of soybean. 35S and NOS represent 35S promoter derived from cauliflower mosaic virus and nopaline synthetase gene terminator, respectively.

Figure 3 shows the construction of expression vectors used for the expression in plants of chimera genes each having a raffinose synthetase gene and a promoter linked thereto. The restriction endonuclease map of the raffinose synthetase gene cloned in the plasmid pBluescriptKS-RS is shown in the upper portion of this figure. pBI121RS and pBI121(-)RS indicate the restriction endonuclease maps of binary vectors used for the transformation of mustard. For the binary vector, only a region between the right border and the left border is shown. 35S, NOS and NPT represent 35S promoter derived from cauliflower mosaic virus, nopaline synthetase gene terminator and kanamycin resistance gene, respectively.

DETAILED DESCRIPTION OF THE INVENTION

The gene engineering methods described below can be carried out according to ordinary methods, for example, as described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press, ISBN 0-87969-309-6; "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X; and "Current Protocols In Protein Science" (1995), John Wiley & Sons, Inc. ISBN 0-471-11184-8.

The term "raffinose synthetase gene" as used herein refers to a gene having a nucleotide sequence coding for the amino acid sequence of a protein capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule (hereinafter referred to simply as the present gene), and such a gene can be prepared, for example, from plants.

More specifically, the present gene can be prepared from dicotyledons such as leguminous plants (e.g., broad bean, soybean) and lamiaceous plants (e.g., Japanese artichoke) or from monocotyledons such as gramineous plants (e.g., corn). Specific examples of the present gene are a "raffinose synth tase gene having a nucleotide sequence cod-

ing for a protein having the amino acid sequence of SEQ ID NO:1"; a "raffinose synthetase gene having a nucleotide sequence coding for a pr t in having an amino acid sequence derived by deletion, replacement, modification or addition of one or sev ral amino acids in the amino acid sequence of SEQ ID NO:1, and capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule"; a "raffinose synthetase gene having a nucleotide sequence coding for a protein having the amino acid sequence of SEQ ID NO:3"; a "raffinose synthetase gene having a nucleotide sequence coding for a protein having an amino acid sequence derived by deletion, replacement, modification or addition of one or several amino acids in the amino acid sequence of SEQ ID NO:3, and capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule"; a "raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of SEQ ID NO:5"; and a "raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of SEQ ID NO:7."

In a further embodiment, the invention relates to nucleic acid molecules of at least 15 nucleotides in length hybridizing specifically with a nucleic acid molecule as described above or with a complementary strand thereof. Specific hybridization occurs preferably under stringent conditions and implies no or very little cross-hybridization with nucleotide sequences encoding no or substantially different (oligo)peptides. Such nucleic acid molecules may be used as probes and/or for the control of gene expression. Nucleic acid probe technology is well known to those skilled in the art who will readily appreciate that such probes may vary in length. Preferred are nucleic acid probes of 17 to 50, particularly preferred 21 to 50 nucleotides in length. Of course, it may also be appropriate to use nucleic acids of up to 100 and more nucleotides in length. The nucleic acid probes of the invention are useful for various applications. On the one hand, they may be used as PCR primers for amplification of regulatory sequences according to the invention. Another application is the use as a hybridization probe to identify nucleic acid molecules hybridizing with a nucleic acid molecule of the invention by homology screening of genomic DNA libraries. Nucleic acid molecules according to this preferred embodiment of the invention which are complementary to a nucleic acid molecule as described above may also be used for repression of expression of a gene comprising such regulatory sequences, for example due to an antisense or triple helix effect or for the construction of appropriate ribozymes (see, e.g., EP-B1 0 291 533, EP-A1 0 321 201, EP-A2 0 360 257) which specifically cleave the (pre)-mRNA of a gene comprising a nucleic acid molecule of the invention. Selection of appropriate target sites and corresponding ribozymes can be done as described for example in Steinecke, Ribozymes, Methods in Cell Biology 50, Galbraith et al. eds Academic Press, Inc. (1995), 449-460. Furthermore, the person skilled in the art is well aware that it is also possible to label such a nucleic acid probe with an appropriate marker for specific applications, such as for the detection of the presence of a nucleic acid molecule of the invention in a sample derived from an organism.

The above described nucleic acid molecules may either be DNA or RNA or a hybrid thereof. Furthermore, said nucleic acid molecule may contain, for example, thioester bonds and/or nucleotide analogues, commonly used in oligonucleotide anti-sense approaches. Said modifications may be useful for the stabilization of the nucleic acid molecule against endo- and/or exonucleases in the cell. Said nucleic acid molecules may be transcribed by an appropriate vector containing a chimeric gene which allows for the transcription of said nucleic acid molecule in the cell. Such nucleic acid molecules may further contain ribozyme sequences which specifically cleave the (pre)-mRNA comprising the nucleic acid molecule of the invention. Furthermore, oligonucleotides can be designed which are complementary to a nucleic acid molecule of the invention (triple helix; see Lee, Nucl. Acids Res. 6 (1979), 3073; Cooney, Science 241 (1988), 456 and Dervan, Science 251 (1991), 1360), thereby preventing transcription and the production of the encoded oligopeptide.

The present gene can be obtained, for example, by the following method.

The tissues of a leguminous plant such as broad bean (*Vicia faba*) or soybean (*Glycine max*) are frozen in liquid nitrogen and ground physically with a mortar or other means into finely powdered tissue debris. From the tissue debris, RNA is extracted by an ordinary method. Commercially available RNA extraction kits can be utilized in the extraction. The whole RNA is separated from the RNA extract by ethanol precipitation. From the whole RNA separated, poly-A tailed RNA is fractionated by an ordinary method. Commercially available oligo-dT columns can be utilized in the fractionation. cDNA is synthesized from the fraction obtained (i.e., poly-A tailed RNA) by an ordinary method. Commercially available cDNA synthesis kits can be utilized in the synthesis.

For example, cDNA fragments of the "raffinose synthetase gene having a nucleotide sequence coding for a protein having the amino acid sequence of SEQ ID NO:1" as the present gene can be obtained by PCR amplification using the broad bean-derived cDNA obtained above as a template and primers 1 to 3 shown in list 1 below. The primers used therein can be designed and synthesized on the basis of the nucleotide sequence of SEQ ID NO:2, depending upon the purpose. For example, in order to amplify the open reading frame region of the "raffinose synthetase gene having a nucleotide sequence coding for a protein having the amino acid sequence of SEQ ID NO:1," primers 1 to 4 shown in list 2 below may be designed and synthesized.

In the same manner, cDNA fragments of th "raffinose synthetase gene having a nucleotide sequence coding for

a protein having the amino acid sequence of SEQ ID NO:3" can be obtained by PCR amplification with the soybean-derived cDNA obtained above as a template and, for example, primers 4 to 6 shown in list 1 below. The primers used ther in can be designed and synthesized on the basis of the nucleotide sequence of SEQ ID NO:4, depending upon the purpose. For example, in order to amplify the open reading frame region of the "raffinose synthetase gene having a nucleotide sequence coding for a protein having the amino acid sequence of SEQ ID NO:3," primers 5 to 8 shown in list 2 below may be designed and synthesized.

The amplified DNA fragments can be subcloned according to ordinary methods, for example, as described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; and "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X. More specifically, cloning can be effected, for example, using a TA cloning kit (Invitrogen) and a plasmid vector such as pBluescript II (Stratagene). The nucleotide sequences of the DNA fragments cloned can be determined by the dideoxy terminating method, for example, as described by F. Sanger, S. Nicklen, A.R. Coulson, Proceedings of National Academy of Science U.S.A. (1977), 74, pp. 5463-5467. For example, ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit commercially available from Perkin-Elmer may preferably be used.

(List 1)

Primer 1: AATTTTCAAG CATAGCCAAG TTAACCACCT 30 mer
Primer 2: GCTCACAAGA TAATGATGTT AGTC 24 mer
Primer 3: ATACAAGTGA GGAACTTGAC CA 22 mer
Primer 4: CCAAACCATA GCAAACCTAA GCAC 24 mer
Primer 5: ACAACAGAAA AATATGACTC TTATTACT 28 mer
Primer 6: AAAAGAGAGT CAAACATCAT AGTATC 26 mer

(List 2)

Primer 1: ATGGCACCAC CAAGCATAAC CAAAACTGC 29 mer
Primer 2: ATGGCACCAC CAAGCATAAC CAAAACTGCA ACCCTCCAAG ACG 43 mer
Primer 3: TCAAAATAAA AACTGGACCA AAGAC 25 mer
Primer 4: TCAAAATAAA AACTGGACCA AAGACAATGT 30 mer
Primer 5: ATGGCTCCAA GCATAAGCAA AACTG 25 mer
Primer 6: ATGGCTCCAA GCATAAGCAA AACTGTGGAA CT 32 mer
Primer 7: TCAAAATAAA AACTCAACCA TTGAC 25 mer
Primer 8: TCAAAATAAA AACTCAACCA TTGACAATTT TGAAGCACT 39 mer

The term "gene fragment" as used herein refers to a gene fragment having a partial nucleotide sequence of the present gene (hereinafter referred to simply as the present gene fragment). For example, it may be a gene fragment derived from a plant and having a partial nucleotide sequence of the gene having a nucleotide sequence coding for a protein capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule. Specific examples of the present gene fragment are a gene fragment having a partial nucleotide sequence of the gene having a nucleotide sequence coding for the amino acid sequence of SEQ ID NO:1 and a gene fragment having a partial nucleotide sequence of the gene having a nucleotide sequence of SEQ ID NO:2, more specifically a gene fragment having a nucleotide sequence or a partial nucleotide sequence thereof, coding for any of the amino acid sequences shown in list 3 below.

These gene fragments can be used as probes in the hybridization method or as primers in the PCR method. For the primers in the PCR method, it is generally preferred that the number of nucleotides is greater from a viewpoint that the specificity of annealing is ensured; it is, however, also preferred that the number of nucleotides is not so great from viewpoints that the primers themselves are liable to have a higher structure giving possible deterioration of the annealing efficiency and that complicated procedures are required in the purification after the synthesis. In usual cases, preferred is a gene fragment consisting of single-stranded DNA, wherein the number of nucleotides is in the range of from 15 to 50.

(List 3)

#1 Gly Ile Lys Phe Met Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp Val Gly
#2 Ile Ile Asp Lys Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr
#3 Gly Gly Cys Pro Pro Gly Phe Val Ile Ile Asp Asp Gly Trp Gln

#4 Thr Ser Ala Gly Glu Gln Met Pro Cys Arg Leu Val Lys Tyr Glu Glu Asn
#5 Val Tyr Val Trp His Ala Leu Cys Gly Tyr Trp Gly Gly Val Arg Pro
#6 Thr Met Glu Asp Leu Ala Val Asp Lys Ile Val Glu Asn Gly Val Gly Leu Val Pro Pro
#7 Gly Leu His Ser His Leu Glu Ser Ala Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu
#8 Gly Gly Arg Val Glu Leu Ala Arg Ala Tyr Tyr Lys Ala Leu
#9 Val Lys Lys His Phe Lys Gly Asn Gly Val Ile Ala
#10 Glu His Cys Asn Asp Phe Phe Leu Leu Gly Thr Glu Ala Ile Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Ser Asp Pro Ser Gly Asp Pro Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val His Cys
#11 Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp
#12 Leu Pro Asp Gly Ser Ile Leu Arg Cys
#13 Ala Leu Pro Thr Arg Asp Cys Leu Phe Glu Asp Pro Leu His Asn Gly Lys Thr Met Leu Lys Ile Trp Asn
#14 Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly Gly Trp
#15 Phe Ala Pro Ile Gly Leu Val Asn Met

The present gene fragment is labeled, and then used as a probe in the hybridization method and hybridized to organism-derived DNA, so that a DNA fragment having the probe specifically bound thereto can be detected. Thus, from an organism-derived gene library, a raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of an enzyme capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule; or a gene fragment having a partial nucleotide sequence thereof, can be detected (hereinafter referred to simply as the present detection method).

As the organism-derived DNA, for example, a cDNA library or a genomic DNA library of a desired plant can be used. The gene library may also be a commercially available gene library as such or a library prepared according to an ordinary library preparation method, for example, as described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor Laboratory Press; "Current Protocols In Molecular Biology" (1987), John Wiley & Sons, Inc. ISBN 0-471-50338-X.

As the hybridization method, plaque hybridization or colony hybridization can be employed, depending upon the kind of vector used in the preparation of a library. More specifically, when a library to be used is constructed with a phage vector, a suitable host microorganism is mixed with the phage under infectible conditions, which is further mixed with a soft agar medium, and the mixture is plated on an agar medium. Thereafter, a culture is grown at 37°C until a plaque of an appropriate size appears. When a library to be used is constructed with a plasmid vector, the plasmid is transformed in a suitable host microorganism to form a transformant. The transformant obtained is diluted to a suitable concentration, and the dilution is plated on an agar medium, after which a culture is grown at 37°C until a colony of an appropriate size appears.

In either case of the above libraries, a membrane filter is placed on the surface of the agar medium after the cultivation, so that the phage or transformant is transferred to the membrane. This membrane is denatured with an alkali, followed by neutralization, and for example, when a nylon membrane is used, the membrane is irradiated with ultraviolet light, so that DNA is fixed on the membrane. This membrane is then subjected to hybridization with the present gene fragment labeled by an ordinary method as a probe. For this method, reference may be made, for example, to D.M. Glover ed., "DNA cloning, a practical approach" IRL PRESS (1985), ISBN 0-947946-18-7. There are various reagents and temperature conditions to be used in the hybridization; for example, prehybridization is carried out by the addition of 6 x SSC (0.9 M NaCl, 0.09 M citric acid), 0.1-1% SDS, 100 μg/ml denatured salmon sperm DNA, and incubation at 65°C for 1 hour. The present gene fragment labeled is then added as a probe, and mixed. Hybridization is carried out at 42-68°C for 4 to 16 hours. The membrane is washed with 2 x SSC, 0.1-1% SDS, further rinsed with 0.2 x SSC, 0-0.1% SDS, and then dried. The membrane is analyzed, for example, by autoradiography or other techniques, to detect the position of the probe on the membrane and thereby detect the position of DNA having a nucleotide sequence homologous to that of the probe used Thus, the present gene or the present gene fragment can be detected. The clone corresponding to the position of DNA thus detected on the membrane is identified on the original agar medium, and the positive clone is selected, so that the clone having the DNA can be isolated. The same procedures of detection are repeated to purify the clone having the DNA.

Other detection methods can also be used, for example, GENE TRAPPER cDNA Positive Selection System Kit commercially available from Gibco BRL. In this method, a single-stranded DNA library is hybridized with the present gene fragment biotinylated (i.e., probe), followed by the addition of streptoavidin-bound magnet beads and mixing. From the mixture, the streptoavidin-bound magnetic beads are collected with a magnet, so that single-stranded DNA having a nucleotide sequence homologous to that of th probe used, which has been bound to these beads through the present gen fragment, biotin and streptoavidin, is collected and detected. Thus, the present gene or th present g ne fragment can be detected. The single-stranded DNA collected can be changed into a double-strand form by treatment

with a suitable DNA polymerase using a suitable oligonucleotide as a primer.

The present detection method may also be used in the analysis of a plant. More specifically, plant genomic DNA is prepared according to an ordinary method, for example, as described in "Cloning and Sequence (Plant Biotechnology Experiment Manual)" complied under the supervision of Itaru Watanabe, edited by Masahiro Sugiura, published by Noson Bunka-sha, Tokyo (1989). The plant genomic DNA is digested with several kinds of suitable restriction endonucleases, followed by electrophoresis, and the electrophoresed DNA is blotted on a filter according to an ordinary method. This filter is subjected to hybridization with a probe prepared from the present gene fragment by an ordinary method, and DNA fragments to which the probe hybridizes are detected. The DNA fragments detected are compared in length between different varieties of a specified plant species. The differences in length make possible the analysis of differences in phenotypic characteristics accompanied with the expression of raffinose family oligosaccharides between these varieties. Furthermore, when the DNA fragments detected by the above method are compared in length between the gene recombinant plant and the non-gene recombinant plant of the same variety, the former plant can be discriminated from the latter plant by the detection of hybridizing bands greater in number or higher in concentration for the former plant than for the latter plant. This method can be carried out according to the RFLP (restriction fragment length polymorphism) method, for example, as described in "Plant PCR Experiment Protocols" complied under the supervision of Ko Shimamoto and Takuji Sasaki, published by Shujun-sha, Tokyo (1995), ISBN 4-87962-144-7, pp. 90-94.

The PCR method using a primer having the nucleotide sequence of the present gene fragment makes it possible to amplify from organism-derived DNA, a raffinose synthetase gene having a nucleotide sequence coding for the amino acid sequence of an enzyme capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule; or a gene fragment having a partial nucleotide sequence thereof (hereinafter referred to simply as the present amplification method).

More specifically, for example, an oligonucleotide having 15 to 50 nucleotides in the nucleotide sequence of the present gene fragment at the 3'-terminus is chemically synthesized by an ordinary synthesis method. Based on the codon table below, showing the correspondence of amino acids encoded in nucleotide sequences, a mixed primer can also be synthesized so that a residue at a specified position in the primer is changed to a mixture of several bases, depending upon the variation of codons which can encode a certain amino acid.

CODON TABLE

Phe	UUU	Ser	UCU	Tyr	UAU	Cys	UGU
	UUC		UCC		UAC		UGC
Leu	UUA		UCA	Stop	UAA	Stop	UGA
	UUG		UCG		UAG	Trp	UGG
	CUU	Pro	CCU	His	CAU	Arg	CGU
	CUC		CCC		CAC		CGC
	CUA		CCA	Gln	CAA		CGA
	CUG		CCG		CAG		CGG
Ile	AUU	Thr	ACU	Asn	AAU	Ser	AGU
	AUC		ACC		AAC		AGC
	AUA		ACA	Lys	AAA	Arg	AGA
Met	AUG		ACG		AAG		AGG
Val	GUU	Ala	GCU	Asp	GAU	Gly	GGU
	GUC		GCC		GAC		GGC
	GUA		GCA	Glu	GAA		GGA
	GUG		GCG		GAG		GGG

Furthermore, a base capable of forming a pair with plural kinds of bases, such as inosine, can also be used instead

of the above mixture of several bases. More specifically, for example, primers having nucleotid sequences as shown in list 4 can be used. In this context, an oligonucleotid having the same nucleotide sequence as th coding strand of the present gene consisting of double-stranded DNA is designated "sense primer," and an oligonucleotide having a nucleotide sequence complementary to the coding strand, "antisense primer."

A sense primer having the same nucleotide sequence as present on the 5'-upstream side in the coding strand of a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof to be amplified, and an antisense primer having a nucleotide sequence complementary to the nucleotide sequence on the 3'-downstream side in this coding strand, are used in combination for PCR reaction to amplify a DNA fragment, for example, with a gene library, genomic DNA or cDNA as a template. At this time, the amplification of a DNA fragment can be confirmed by an ordinary method with electrophoresis. For the DNA fragment amplified, its restriction endonuclease map is constructed or its nucleotide sequence is determined by an ordinary method, so that the present gene or the present gene fragment can be identified. As the gene library used herein, for example, a cDNA library or a genomic cDNA library of a desired plant can be used. For the plant gene library, a commercially available library derived from plant can be used as such; or a library prepared according to an ordinary library preparation method, for example, as described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor laboratory Press or "Current Protocols in Molecular Biology" (1987), John Wiley & Sons, Inc., ISBN 0-471-50338-X, can also be used. As the genomic DNA or cDNA used in the present amplification method, for example, cDNA or genomic cDNA prepared from a desired plant can be used.

More specifically, for example, a primer designed on the amino acid sequence of SEQ ID NO:1 is used for the present amplification method with cDNA derived from Japanese artichoke, which is a lamiaceous plant, as a template, so that a raffinose synthetase gene fragment having the nucleotide sequence of SEQ ID NO:6 can be amplified. Furthermore, for example, a primer designed on the amino acid sequence of SEQ ID NO:1 is used for the present amplification method with cDNA derived from corn, which is a gramineous plant, as a template, so that a raffinose synthetase gene fragment having the nucleotide sequence of SEQ ID NO:8 can be amplified.

(List 4)

1-F	32mer	
	TTIAAIGTITGGTGGACIACICAITGGGTIGG	
2-F	41mer	
	ATIATIGAIAAITTIGGITGGTGIACITGGGAIGCITTITA	
2-RV	41mer	
	TAIAAIGCITCCCAIGTICACCAICCIAAITTITCIATIAT	
3-F	44mer	
	GGIGGITGICCICCIGGITTIGTIATIATIGAIGAIGGITGGCA	
3-RV	44mer	
	TGCCAICCITCITCIATIATIACIAAICCIGGIGGICAICCICC	
4-F	32mer	
	AAIAAICAITTIAAIGGIAAIGGIGTIATIGC	
4-RV	32mer	
	GCIATIACICCITTICCITTIAAITGITTITT	
5-F	38mer	
	TGGATGGGIAAITTIATICAICCIGAITGGGAIATGTT	
5-RV	38mer	
	AACATITCCCAITCIGGITGIATIAAITTICCCATCCA	
6-RV	27mer	
	CATITTIACIA(AG)ICCIATIGGIGCIAA	

The present amplification method can also be utilized for the analysis of a plant gene. More specifically, for example, plant genomic DNA prepared from different vaneties of a specified plant species is used as a template for the present amplification method to amplify a DNA fragment. The DNA fragment amplified is mixed with a solution of formaldehyde, which is denatured by heating at 85°C for 5 minutes, followed by rapid cooling on ice. This sample is subjected to electrophoresis, for example, on a 6% polyacrylamide gel containing 0% or 10% glycerol. In this electrophoresis, a commercially available electrophoresis apparatus for SSCP (single strand conformation polymorphism) can be used, and electrophoresis is carried out, while the gel is kept at a constant temperature, e.g., 5°C, 25°C or 37°C. From the electrophoresed gel, a DNA fragment is detected, for example, by a method such as silver staining method with commercially available reagents.

From th diff rences of behavior betwe n the varieties in the lectrophoresis of the DNA fragment detected, a mutation in the raffinos synthetase gen is detected, and an analysis is carried out for differences caused by the muta-

tion in phenotypic characteristics accompanied with the expressi n of raffinose family oligosaccharides. This method can be carried out according to the SSCP method, for example, as described in "Plant PCR Experiment Protocols" complied under the supervision of Ko Shimamoto and Takuji Sasaki, published by Shujun-sha, Tokyo (1995), ISBN 4-87962-144-7, pp. 141-146.

The present detection method or the present amplification method as described above can also be used for identifying a raffinose synthetase gene or a gene fragment having a partial nucleotide sequence thereof and then isolating and purifying the identified gene or gene fragment thereof to obtain the present gene (hereinafter referred to simply as the present gene acquisition method).

The present gene or the present gene fragment can be obtained, for example, by detecting a probe consisting of the present gene fragment hybridized to DNA in the organism-derived gene library by the present detection method as described above to identify DNA having a nucleotide sequence homologous with the probe used; purifying a clone carrying the DNA; and isolating and purifying plasmid or phage DNA from the clone. When the DNA thus obtained is a gene fragment having a partial nucleotide sequence of the raffinose synthetase gene, further screening of the gene library by the present gene detection method using the DNA as a probe gives the present gene in full length.

The present gene or the present gene fragment can be identified, for example, by effecting polymerase chain reaction using a primer having the nucleotide sequence of the present gene fragment to amplify a DNA fragment from the organism-derived DNA by the present amplification method as described above; and then constructing a restriction endonuclease map or determining a nucleotide sequence for the amplified DNA fragment. Based on the nucleotide sequence of the gene fragment obtained, an antisense primer is synthesized for the analysis of 5'-terminal sequences, and a sense primer is synthesized for the analysis of 3'-terminal sequences. The nucleotide sequence of the present gene in full length can be determined by the RACE method using these primers and a commercially available kit, e.g., Marathon Kit of Clontech. The present gene in full length can be obtained by synthesizing new primers based on both terminal sequences in the nucleotide sequence thus determined and effecting polymerase chain reaction again.

The present gene acquisition method as described above makes it possible to obtain raffinose synthetase genes as the present gene from various organisms. For example, a gene coding for a raffinose synthetase having an amino acid sequence that has about 50% or higher homology, in the region corresponding to the length of 400 or more amino acids, with the amino acid sequence of SEQ ID NO:1, and capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule. More specifically, for example, a raffinose synthetase gene having the nucleotide sequence of SEQ ID NO:4 can be obtained by amplifying and identifying a DNA fragment containing a gene fragment having a partial nucleotide sequence of the raffinose synthetase gene by the present amplification method using primers designed from the amino acid sequence of SEQ ID NO:1 and soybean cDNA as a template; isolating and purifying the identified DNA fragment, followed by the above procedures to obtain a full-length gene containing the DNA fragment.

A chimera gene comprising the present gene and a promoter linked thereto (hereinafter referred to simply as the present chimera gene) can be constructed.

The promoter to be used is not particularly limited, so long as it is functionable in a host organism to be transformed. The promoter may include, for example, synthetic promoters functionable in *Escherichia coli*, such as *E. coli* lactose operon promoter, *E. coli* tryptophan operon promoter and tac promoter; yeast alcohol dehydrogenase gene (ADH) promoter, adenovirus major late (Ad.ML) promoter, SV40 early promoter, and baculovirus promoter.

When the host organism is a plant or a cell thereof, the promoter may include, for example, T-DNA derived constitutive promoters such as nopaline synthetase gene (NOS) promoter and octopine synthase gene (OCS) promoter; plant virus-derived promoters such as cauliflower mosaic virus (CaMV) derived 19S and 35S promoter; derived promoters such as phenylalanine ammonia-lyase (PAL) gene promoter, chalcone synthetase (CHS) gene promoter and pathogenesis-related protein (PR) gene promoter. Furthermore, vector pSUM-GY1 (see JP-A 06-189777/1994) can also be used, which has a promoter giving specific expression in a specified plant tissue, e.g., a promoter of soybean-derived seed storage protein glycinin gene. The use of a chimera gene constructed so as to have such a promoter makes it possible to increase or decrease the content of raffinose family oligosaccharides in a specified tissue of a plant.

The present chimera gene is then introduced into a host organism according to an ordinary gene engineering method to give a transformant. If necessary, the present chimera gene may be used in the form of a plasmid, depending upon the transformation method for introducing the gene into the host organism. Furthermore, the present chimera gene may contain a terminator. In this case, it is generally preferred that the chimera gene is constructed so as to have a terminator downstream the raffinose synthetase gene. The terminator to be used is not particularly limited, so long as it is functionable in a host organism to be transformed. For example, when the host organism is a plant or a cell thereof, the terminator may include, for example, T-DNA derived constitutive terminators such as nopaline synthetase gene (NOS) terminator; and plant derived terminators such as terminators of allium virus GV1 or GV2.

If necessary, the present gen may be used in the form of a plasmid. For example, when the host organism is a microorganism, the plasmid constructed is introduced into the microorganism by an ordinary means, for exampl , as described in "Molecular Cloning: A Laboratory Manual 2nd edition" (1989), Cold Spring Harbor laboratory Press or

"Current Protocols in Molecular Biology" (1987), John Wiley & Sons, Inc., ISBN 0-471-50338-X. The microorganism thus transformed is selected with a marker such as antibiotic resistanc or auxotrophy. When the host organism is a plant, the plasmid constructed is introduced into a plant cell by an ordinary means such as infection with Agrobacterium (see JP-B 2-58917/1990 and JP-A 60-70080/1985), electroporation into protoplasts (see JP-A 60-251887/1985 and JP-B 5-68575/1993) or particle gun method (see JP-A 5-508316/1993 and JP-A 63-258525/1988). The plant cell transformed by the introduction of a plasmid is selected with an antibiotic such as kanamycin or hygromycin. From the plant cell thus transformed, a transformed plant can be regenerated by an ordinary plant cell cultivation method, for example, as described in "Plant Gene Manipulation Manual (How to Produce Transgenic Plants)" written by Uchimiya, 1990, Kodan-sha Scientific (ISBN 4-06-153513-7), pp. 27-55. Furthermore, the collection of seeds from the transformed plant also makes it possible to prolify the transformed plant. In addition, crossing between the transformed plant obtained and the non-transformed plant makes it possible to produce progenic plants with the character of the transformed plant.

The content of raffinose family oligosaccharides can be changed by introducing the present gene into a host organism or a cell thereof, and modifying the metabolism in the host organism or the cell thereof. As such a method, for example, there can be used a method for metabolic modification to increase the amount of raffinose family oligosaccharides in a host organism or a cell thereof by constructing the present chimera gene comprising the present gene and a promoter linked thereto, in which case the present gene is linked to the promoter in an original direction suitable for transcription, translation, and expression as a protein, and then introducing the present chimera gene into the host organism or the cell thereof; or a method for metabolic modification to decrease the amount of raffinose family oligosaccharides in a host organism or a cell thereof by constructing the present chimera gene comprising the present gene and a promoter linked thereto, in which case the present gene is linked to a promoter in a reverse direction unsuitable for translation and expression as a protein, and then introducing the present chimera gene into the host organism or the cell thereof.

The term "raffinose synthetase protein" as used herein refers to a protein encoded in the present gene (hereinafter referred to simply to the present protein). For example, it may include an enzyme protein having the amino acid sequence of SEQ ID NO:1 or SEQ ID NO: 3, or having an amino acid sequence derived by deletion, replacement, modification or addition of one or several amino acids in the amino acid sequence of SEQ ID NO:1 or SEQ ID NO:3; and capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule.

Specific examples of the present protein are an enzyme protein having the amino acid sequence of SEQ ID NO:1 (799 amino acids; molecular weight, 89 kDa) and an enzyme protein having the amino acid of SEQ ID NO:3 (781 amino acids; molecular weight, 87 kDa).

The present protein, although it can be prepared, for example, from leguminous plants such as broad bean (*Vicia faba*), by an ordinary biochemical method such as $(NH_4)_2SO_4$ precipitation, ion exchange column, hydrophobic column, hydroxyapatite column or gel filtration column, can also be prepared from the host organism transformed with the present plasmid, or a cell thereof. More specifically; for example, using GST Gene Fusion Vectors Kit of Pharmacia, the present gene is inserted into an expression vector plasmid attached to the Kit. The resulting vector plasmid is introduced into a microorganism such as *E. coli* according to an ordinary gene engineering method. A culture of the transformant obtained is grown on a medium with the addition of IPTG (isopropylthio-β-D-galactoside), so that the present protein can be expressed and derived as a fused protein in the culture. The fused protein expressed and induced can be isolated and purified by an ordinary method such as disruption of bacterial cells, column operation or SDS-PAGE electrophoresis. The digestion of the fused protein with a protease such as thrombin or blood coagulation factor Xa gives the present protein. This may preferably be made, for example, according to the method described in "Current Protocols In Protein Science" (1995), John Wiley & Sons, Inc. ISBN 0-471-11184-8. The activity of the present protein can be measured, for example, by the method described in L. Lehle and W. Tanner, Eur. J. Biochem., 38, 103-110 (1973).

An anti-raffinose synthetase antibody capable of binding to a raffinose synthetase protein (hereinafter referred to simply as the present antibody) can be produced by an ordinary immunological method using the present protein prepared above, as an antigen. More specifically, the present antibody can be produced, for example, according to the method described in Ed Harlow and David Lane, "Antibodies: A Laboratory Manual" (1988), Cold Spring Harbor Laboratory Press, ISBN 0-87969-314-2.

The present protein can be detected by treating test proteins with the present antibody and detecting a protein having the present antibody bound specifically thereto. Such a detection method can be carried out according to an immunological technique such as Western blot method or enzyme-linked immunosorbent assay (ELISA), for example, as described in Ed Harlow and David Lane, "Antibodies: A Laboratory Manual" (1988), Cold Spring Harbor Laboratory Press.

Th Western blot method is carried out, for example, as follows: Proteins are extracted from a plant, for example, according to the method described in Methods in Enzymology, volume 182, "Guide to Protein Purification," pp. 174-193, ISBN 9-12-182083-1. The composition of an extraction buffer can suitably be changed depending upon the plant tissue

used. The proteins extracted are electrophoresed according to an ordinary SDS-PAGE method. The proteins lectrophoresed in the gel ar transferred to a membran by Western blotting with an ordinary electrical method. Mor specifically, for example, the gel is immersed in a transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) for 10 minutes, and then placed onto a PVDF membrane cut into the same size as that of the gel. The gel together with the membrane is set in a commercially available transfer apparatus of the semi-dry type. Blotting is carried out at a constant current of 0.8 to 2 mA/cm^2 for 45 minutes to 1 hour. The proteins transferred to the membrane can be detected immunologically with a kit for Western blot detection using a primary antibody, and a secondary antibody or protein A, which has been labeled with alkaline phosphatase or horseradish peroxidase. At this time, the present protein on the membrane call be detected by the use of the present antibody as a primary antibody.

In the ELISA method, for example, the property of proteins binding to the surface of a 96-well ELISA plate made of a resin is utilized in principle for the immunological detection of an antigen finally bound to the surface of the ELISA plate. The test proteins are added as a solution and bound to an ELISA plate, followed by blocking, for example, by the addition of PBS containing a protein such as 5% bovine serum albumin. Thereafter, the well is washed with PBS, to which a solution containing the present antibody is added to effect the reaction. After the well is washed, a solution containing a secondary antibody labeled with alkaline phosphatase or horseradish peroxidase is further added to the well, followed by washing. Finally, a substrate solution for detection is added to the well, and the color development of the substrate is detected with an ELISA reader.

In another method, the present antibody is added and bound to an ELISA plate, followed by blocking, for example, by the addition of PBS containing a protein such as 5% bovine serum albumin. The test proteins are then added as a solution, and an antigen contained in the test proteins is bound to the present antibody that has been bound to the plate, followed by washing, and the present antibody is further added to the well. The present antibody used at this time is preferably one prepared from an animal species different from that used for the preparation of the present antibody used first. A solution containing a secondary antibody labeled with alkaline phosphatase or horseradish peroxidase is then added to the well, followed by washing. The secondary antibody used at this time must have the property of binding to the present antibody added later. Finally, a substrate solution for detection is added, and the color development is detected with an ELISA reader.

Examples

The present invention will be further illustrated by the following examples; however, the present invention is not limited to these example in any way whatsoever.

Example 1 (Purification of Galactinol)

About 250 ml of sugar beet blackstrap molasses was five-fold diluted with methanol. The dilution was centrifuged at 21,400 x g for 15 minutes at room temperature to remove insoluble matter. The supernatant obtained was transferred into a 2-liter Erlenmeyer flask, to which isopropanol at a half volume was added portionwise with stirring. The flask was left at room temperature for a while until the resulting precipitate adhered to the wall of the flask. The supernatant was then discarded by decantation. To the precipitate was added 500 ml of ethanol, and the mixture was washed by stirring with a rotary shaker. The washing was further repeated several times. The washed precipitate was scraped off from the wall of the flask, followed by air drying on a filter paper. The air-dried precipitate (dry powder) was dissolved in purified water to about 40% (w/v). To this solution was added AG501-X8(D) of BioRad, followed by stirring. This operation was repeated until the color of the solution became almost unobserved. The resulting solution was treated with a Sep-Pak QMA column of Millipore, and further pretreated with Sep-Pak CM, Sep-Pak C18 and Sep-Pak Silica columns of Millipore. The resulting solution was loaded at a volume of 5 ml onto a column of Wako-gel LP40C18 (Wako Pure Chemical Industries, 2.6 cm x 85 cm), and eluted with purified water. The sugar content of the eluate was measured with a portable sugar refractometer, and the sugar composition was analyzed by high performance liquid chromatography (HPLC) with a Sugar-pak Na (7.8 mm x 300 mm) column of Millipore. The detection of sugars was carried out with model 410 Differential Refractometer of Waters. The eluate containing galactinol was lyophilized, and the resulting lyophilized powder was dissolved in 5 ml of purified water. The solution was loaded onto a column of TOYOPEARL HW40(S) (Toso, 2.6 cm x 90 cm), and eluted with purified water. The eluate was analyzed in the same manner as described above, so that purified galactinol was obtained.

The galactinol obtained was kept at 25°C for 40 minutes in the reaction mixture that came to contain 80 mM phosphate buffer (pH 6.5), 2 mg/ml galactinol, and 8.3 U α-galactosidase (Boehringer Mannheim, *E. coli* overproducer 662038). The reaction mixture was extracted with chloroform, and the water layer was analyzed by HPLC. The resulting galactinol was confirmed to be hydrolyzed into galactose and myo-inositol.

Example 2 (Measurement of Raffinose Synthase Activity)

The raffinose synthetas activity was measured under the following conditions according to the description of L. Lehle and W. Tanner, Eur. J. Biochem., 38, 103-110 (1973).

First, 2 μl of a sample to be used in the measurement of activity was added to 18 μl of the reaction mixture that came to contain 100 mM Tris-HCl (pH 7.4), 5 mM DTT (dithiothreitol), 0.01% BSA, 200 μM sucrose, 5 mM galactinol, 740 KBq/ml (31.7 μM) [^{14}C] sucrose, and the reaction mixture was kept at 37°C for 3 to 20 hours. After the reaction, 30 μl of ethanol was added to the reaction mixture, followed by stirring and centrifugation at 15,000 rpm for 5 minutes. The supernatant was spotted at a volume of 5 μl on an HPTLC plate of cellulose for thin layer chromatography (Merck, 10 cm x 20 cm), and developed with n-butanol : pyridine : water : acetic acid = 60 : 40 : 30 : 3. The developed plate was dried and then quantitatively analyzed with an imaging analyzer (Fuji Photographic Film, FUJIX Bio Imaging Analyzer BAS-2000II) for the determination of [^{14}C] raffinose produced.

Example 3 (Purification of Raffinose Synthase)

The purification of raffinose synthetase from broad bean was carried out as follows: For each purified protein solution, proteins present in the protein solution were analyzed by SDS-PAGE (Daiichi Kagaku Yakuhin), and the enzyme activity thereof was measured according to the method described in Example 2.

First, 300 g of immature seeds of broad bean (Nintoku Issun) stored at -80°C was thawed and then peeled. The peeled seeds were put in 600 ml of 100 mM Tris-HCl (pH 7.4), 5 mM DTT (dithiothreitol), 1 mM EDTA, 1 mM PMSF (phenylmethylsulfonyl fluoride) and 1 mM benzamide, and ground on ice with a mortar. The ground material was centrifuged at 21,400 x g for 50 minutes at 4°C. To the resulting supernatant was added 10% polyethylene imine (pH 8.0) at a 1/20 volume. The mixture was stirred at 4°C for 15 minutes, and centrifuged at 15,700 x g for 20 minutes at 4°C. To the resulting supernatant was added 196 g/l of $(NH_4)_2SO_4$ with stirring. The mixture was stirred in ice for 30 minutes, and centrifuged at 15,700 x g for 20 minutes at 4°C. To the resulting supernatant was further added 142 g/l of $(NH_4)_2SO_4$ with stirring. After the stirring in ice for 30 minutes, the mixture was centrifuged at 15,700 x g for 20 minutes at 4°C. The resulting precipitate was dissolved in 50 ml of 100 mM Tris-HCl (pH 7.4) and 5 mM DTT (dithiothreitol), and the solution was dialyzed against 20 mM Tris-HCl (pH 7.4), 1 mM DTT (dithiothreitol) and 1 mM EDTA at 4°C overnight. After the dialysis, the suspension was centrifuged at 70,000 x g for 60 minutes at 4°C. To the resulting supernatant was added 1 mM benzamidine • HCl, 5 mM ε-amino-n-caproic acid, 1 μg/ml antipain, 1 μg/ml leupeptin and 10 mM EGTA, and 2 M KCl was further added portionwise at a 1/40 volume. The mixture was loaded onto a column of DEAE-Sephacel (Pharmacia, 2.5 cm x 21.5 cm) equilibrated with 0.05 M KCl, 20 mM Tris-HCl (pH 7.4), 1 mM DTT (dithiothreitol) and 1 mM EDTA, and the adsorbed proteins were eluted with a gradient of 0.05 to 0.5 M KCl. The purification steps up to this stage were repeated three times, and fractions having raffinose synthase activity were combined and then purified as follows:

To the eluted fraction having raffinose synthetase activity was added portionwise saturated $(NH_4)_2SO_4$ at a 1/4 volume. The solution was loaded onto a column of Phenyl-Sepharose (Pharmacia, 2.5 cm x 10.2 cm) equilibrated with 20% saturated $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 7.4), 1 mM DTT (dithiothreitol) and 1 mM EDTA, and the adsorbed proteins were eluted with a gradient of 20% to 0% $(NH_4)_2SO_4$. The resulting active fraction was diluted by the addition of 0.01 M potassium phosphate buffer (pH 7.5) at a 2-fold volume. The diluted solution was loaded onto a column of Econo-Pac 10DG (BioRad, 5 ml) previously equilibrated with 0.01 M potassium phosphate buffer (pH 7.5) and 2 mM DTT (dithiothreitol), and the adsorbed proteins were eluted with a gradient of 0.01 to 0.5 M potassium phosphate buffer (pH 7.5) and 2 mM DTT (dithiothreitol). The active fraction obtained at this stage was found to have been purified up to 6500-fold or higher specific activity. Part of the resulting purified protein solution having raffinose synthetase activity was loaded onto a column of Superdex 200 (Pharmacia, 1.6 cm x 60 cm) equilibrated with 0.2 M KCl, 20 mM Tris-HCl (pH 7.4), 1 mM DTT (dithiothreitol) and 1 mM EDTA. The purified proteins thus separated were subjected to SDS-PAGE, and the raffinose synthetase activity was measured. A protein band having raffinose synthetase activity was identified as having a molecular weight of about 90 kDa on the SDS-PAGE.

Example 4 (Analysis of Partial Amino Acid Sequence of Raffinose Synthase)

To about 1 ml of the purified protein solution, which had been purified with a column of Econo-Pac 10DG (BioRad, 5 ml) in Example 3, was added 100% TCA at a 1/9 volume, and the mixture was left on ice for 30 minutes. After centrifugation at 10,000 x g for 15 minutes, the resulting precipitate was suspended in 500 μl of cold acetone (-20°C), followed by further centrifugation. This acetone washing was repeated, and the collected precipitate was dried and then dissolved in 200 μl of SDS-sample buffer, followed by SDS-PAGE. CBB staining was effected for the electrophoresed gel, from which the band of a raffinose synthetase protein was cut out.

To the gel thus taken was added 1 ml of 50% acetonitrile and 0.2 M ammonium carbonate (pH 8.9), and washing

was continued with stirring at room temperature for 20 minutes. The gel was washed once again in the same mann r, and dried under reduced pressure to an extent giving a volum reduction. To this gel was 1 ml of 0.02% Tween-20 and 0.2 M ammonium carbonate (pH 8.9), and the mixture was stirred at room temperature for 15 minutes. After removal of th solution, 400 μl of 8 M urea and 0.4 M NH_4HCO_3 was added, to which 40 μl of 45 mM DTT (dithiothreitol) was further added, and the mixture was left at 50°C for 20 minutes. After complete return to room temperature, 4 μl of 1 M iodoacetic acid was added, and the mixture was stirred in the dark at room temperature for 20 minutes. After removal of the solution, 1 ml of purified water was added, and the mixture was stirred at room temperature for 5 minutes, followed by washing. After further two washings, 1 ml of 50% acetonitrile and 0.2 M ammonium carbonate (pH 8.9) was added, and the mixture was stirred at room temperature for 15 minutes. The same treatment was repeated once again, after which the solution was removed, and the gel was dried under reduced pressure to an extent giving a volume reduction.

To this gel was added a solution of Achromobacter Protease I (Takara, Residue-specific Protease Kit) at a volume of 100 μl. Further added was 0.02% Tween-20 and 0.2 M ammonium carbonate (pH 8.9) to an extent that the gel was not exposed from the surface of the solution, and the mixture was left at 37°C for 42 hours. Further added was 500 μl of 0.09% TFA and 70% acetonitrile, and the mixture was stirred at room temperature for 30 minutes. The resulting mixture as contained in a sample tube was floated in an ultrasonic bath, followed by ultrasonic treatment (BRANSON, 60 W output power) for 5 minutes. The tube and contents thus treated were centrifuged, and the resulting extract was collected in another silicone-coated sample tube. On the other hand, 500 μl of 0.09% TFA and 70% acetonitrile was added again to the precipitate, followed by repeated extraction in the same manner as described above. The resulting extracts were combined and then concentrated under reduced pressure to an extent giving a solution remained at a volume of 200 to 300 μl. To the concentrate was added 25 μl of 8 M urea and 0.4 M NH_4HCO_3, and the mixture was concentrated to an extent giving a solution remained at a volume of 100 μl or lower. The concentrate was brought to about 100 μl with purified water, and the mixture was filtered through a filter of Ultrafree C3 GV (Millipore). The filtrate obtained was then subjected to elution through a column of Aquapore BU-300 C-4 (2.1 mm x 300 mm) by a gradient of 0.1% TFA/2.1% to 68.6% acetonitrile. Absorbance at 215 nm was monitored to collect a fraction at a peak thereof. The sample collected was evaporated under reduced pressure to complete dryness, and then analyzed with a Protein Sequencer 473A of ABI to determine a partial amino acid sequence of a raffinose synthetase.

Example 5 (Preparation of cDNA)

About 2 g of immature seeds of broad bean (Nintoku Issun) was frozen in liquid nitrogen and then ground with a mortar, to which 20 ml of Isogen (Nippon Gene) was added, and the mixture was further thoroughly ground. The ground material was transferred into a centrifugation tube, to which 4 ml of chloroform was added, and the mixture was stirred with a vortex mixer and then centrifuged at 6,500 x g for 10 minutes at 4°C. The water layer was collected, to which 10 ml of isopropanol was added, and the mixture was stirred and then centrifuged at 6,500 x g for 10 minutes at 4°C. The resulting precipitate was washed with 10 ml of 70% ethanol and then dissolved in 1 ml of elution buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1% SDS). The solution was left at 60°C for 10 minutes and then centrifuged at 10,000 x g for 1 minute to remove insoluble matter. To the resulting supernatant was added an equivalent volume of Oligotex-dT30 (Takara), and the mixture was stirred and then left at 65°C for 5 minutes. The mixture was placed on ice and then left for 3 minutes, to which 200 μl of 5 M NaCl was added, and the mixture was left at 37°C for 10 minutes. The mixture was then centrifuged at 10,000 x g at 4°C for 3 minutes. The precipitate was collected and then suspended in 1 ml of TE buffer, and the suspension was left at 65°C for 5 minutes, which was placed on ice and then left for 3 minutes, followed by centrifugation at 10,000 x g for 3 minutes at 4°C to remove the precipitate.

To the resulting supernatant were added 100 μl of 3 M sodium acetate and 2 ml of ethanol, and RNA was ethanol precipitated and collected. The collected RNA was washed twice with 70% ethanol and then dissolved in 20 μl of sterilized water, which was used for the subsequent cDNA synthesis. The amount of RNA obtained was determined by the measurement of absorbance at 260 nm.

For the cDNA synthesis, First Strand Synthesis Kit for RT-PCR (Amercham) and cDNA Synthesis Kit (Takara) were used, and all operations were made according to the protocol.

Example 6 (Nucleotide sequence Analysis of Raffinose Synthase Gene from cDNA)

Based on the amino acid sequence obtained in Example 4, mixed synthetic DNA primers having the nucleotide sequences shown in list 5 below were synthesized. The PCR method was carried out with Gene Amp PCR Systems 2400 and DNA Thermal Cycler Model 480 of Perkin-Elmer using Advantage KlenTaq cDNA Kit of Clontech. The polymerase chain reaction was effected with the above primers at 94°C for 1 minute, at 50°C for 3 minutes, and at 72°C for 3 minutes, and this reaction was repeated forty times. As a result, the combinations of primers 8.2 and 13.3RV, primers 13.4 and 10.3RV, and primers 7.4 and 10.3RV, having the nucleotide sequences shown in list 5 below, gave an

amplification of 1.2 kb, 0.5 kb, and 1.2 kb bands, respectively. Thes amplified DNA fragments were cloned with a TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elm r and nucleotide sequence analysis with a 373S DNA sequencer of ABI. As a result, these DNA fragments were found to have a nucleotide sequence extending from base 813 to base 1915, base 1936 to base 2413, and base 1226 to base 2413, respectively, in the nucleotide sequence of SEQ ID NO:2. Based on these nucleotide sequences, synthetic DNA primers having nucleotide sequences shown in list 6 below were prepared, and the nucleotide sequences in both terminal regions of cDNA were analyzed with Marathon cDNA Amplification Kit of Clontech. As a result, the nucleotide sequence of SEQ ID NO:2 was finally determined.

(List 5)

#8.2	26mer
	AA(AG) AC(ATGC) GC(ATGC) CC(ATGC) AG(TC) AT(TCA) AT(TCA) GAC AA
#13.4	20mer
	AA(AG) AT(TCA) TGG AA(TC) CT(ATGC) AAC AA
#7.4	24mer
	AA(AG) GC(ATGC) AG(AG) GT(ATGC) GT(ATGC) GT(ATGC) CC(ATGC) AAG
#13.3RV	21mer
	(TC)TT (AG)TT (ATGC)AG (AG)TT CCA (AGT)AT TTT
#10.3RV	21mer
	(TC)TT (AG)TC (TC)TC (AG)TA (ATGC)AG (AG)AA TTT

(List 6)

RS-2RV	30mer
	GGCTGAGGTTCGGTTCATTCCTGAATCATC
RS-7	30mer
	CCAAATGGTACATATTGGCTCCAAGGTTGT
RS-8	30mer
	AAGAGTGTATCTGAATTTTCACGCGCGGTG
RS-9	30mer
	TGGTGCAATGGGAAAACTCCAATGAGCACC
RS-10	30mer
	ATGAAGTGTTCTGATAGATTGAAAGTTTCG
RS-11	30mer
	CAGTCTCTGGAGTTTGATGATAATGCAAGT

Example 7 (Cloning of Raffinose Synthetase Gene from Broad Bean cDNA)

The primers designed from the amino acid sequence of SEQ ID NO:1, i.e., primers having nucleotide sequences shown in list 7 below, were synthesized. Using these primers and cDNA obtained in Example 5 as a template, a DNA fragment of the open reading frame region was amplified by PCR under the conditions described in Example 6. The amplified DNA fragment was digested with the restriction endonucleases *Bam* HI and *Xba* I whose recognition sequences were contained in the primers used. Using Ligation Kit (Takara), the DNA fragment thus digested was cloned in the plasmid pBluescriptII KS- (Stratagene) previously digested with *Bam* HI and *Xba* I. The nucleotide sequence of the cloned DNA fragment was confirmed with ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer. In the clone thus obtained, it was found that the base at position 1591 in the nucleotide sequence of SEQ ID NO:2 had been changed from thymine (T) to cytosine (C). This was, however, a nonsense mutation without a change of the amino acid; therefore, this clone was designated pBluescriptKS-RS, and used in the subsequent experiment.

(List 7)

RS-N	41mer
	CGCGGATCCACCATGGCACCACCAAGCATAACCAAAACTGC
RS-C	37mer
	TGCTCTAGATTATCAAAATAAAAACTGGACCAAAGAC

Example 8 (Expression f Broad Bean Raffinose Synthetase Gene in *E. coli*)

The plasmid pBluescriptKS-RS having the broad bean raffinose synthetase gene obtained in Example 7 was digested with *Bam* HI and *Not* I, and cloned in the plasmid pGEX-4T3 (Pharmacia) digested with *Bam* HI and *Not* I to give the plasmid pGEX-RS as shown in Figure 1.

The plasmid pBluescriptKS-RS was digested with *Nco* I and *Xba* I, and cloned in the plasmid pTrc99A (Pharmacia) digested with *Nco* I and *Xba* I to give the plasmid pTrc-RS as shown in Figure 1.

These plasmids were introduced into *E. coli* strain HB101, and the resulting transformants were used for the confirmation of raffinose synthetase expression. Overnight cultures of the transformants were inoculated at a volume of 1 ml each into 100 ml of LB medium and incubated at 37°C for about 3 hours, followed by the addition of IPTG (isopropylthio-β-D-galactoside) to a final concentration of 1 mM and further incubation for 5 hours. The cultures were centrifuged at 21,400 x g for 10 minutes, and the bacterial cells were collected. The collected bacterial cells were stored at -80°C. To the frozen bacterial cells was added a 10-fold volume of 100 mM Tris-HCl (pH 7.4), 1 mM EDTA, 5 mM DTT (dithiothreitol), 1 mM PMSF (phenylmethylsulfonyl fluoride) and 1 mM benzamide, and the bacterial cells were thawed and suspended. These suspensions were treated with an ultrasonic disrupter (Branson) to effect the disruption of the bacterial cells. The disrupted cell mixtures obtained were centrifuged at 16,000 x g for 10 minutes, and soluble protein solutions were collected.

The protein solutions thus obtained were used at a volume of 4 μl each for the measurement of raffinose synthetase activity according to the method described above. The reaction was effected at 37°C for 64 hours. As a control, *E. coli* strain HB101 that had been transformed with one of the vectors, pGEX-4T3, was used. The results are shown in Table 1. The synthesis of raffinose was detected in the samples from the transformants HB101 (pGEX-RS) and HB101 (pTrc-RS).

TABLE 1

Transformant	Amount of raffinose produced (pmol)
HB101 (pGEX4T-3)	0.56
HB101 (pGEX-RS)	10.50
HB101 (pTrc-RS)	11.10

Example 9 (Cloning of Raffinose Synthetase Gene from Soybean cDNA)

In the same manner as described in Example 5, cDNA was obtained from immature seeds of soybean (*Glycine max*) Williams 82. Using this cDNA as a template and primers designed from the amino acid sequence of SEQ ID NO:1, i.e., primers having nucleotide sequences shown in list 8 below, a DNA fragment was amplified by PCR under the conditions described in Example 6. The DNA fragment thus amplified by PCR was cloned with a TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and nucleotide sequence analysis with a 373S DNA sequencer of ABI. Based on this sequence, primers having nucleotide sequences shown in list 9 below were synthesized. The synthesis of cDNA was carried out with Marathon Kit of Clontech using mRNA obtained in the same manner as described in Example 5 from leaves of soybean Williams 82. The cDNA obtained was ligated to an adaptor contained in this kit with ligase. This operation was made according to the protocol attached. Using the adaptor-ligated cDNA thus prepared, polymerase chain reaction was effected with the primers shown in list 9 below. The nucleotide sequences in both terminal regions of the gene were analyzed according to the protocol attached to the Marathon Kit of Clontech. As a result, the nucleotide sequence of SEQ ID NO:4 was determined.

(List 8)

1-F primer	35mer
	CGATTIAAIGTITGGTGGACIACICAITGGGTIGG
2-RV primer	45mer
	GGCCTAIAAIGCITCCCAIGTICACCAICCIAAITTITCIATIAT
5-F primer	41mer
	CGATGGATGGGIAAITTIATICAICCIGAITGGGAIATGTT
6-RV primer	32mer

GGCCACATITTIACIA(AG)ICCIATIGGIGCIAA

(List 9)

SN-1	30mer	
	CACGAACTGGGGCACGAGACACAGATGATG	
SC-3RV	30mer	
	AAGCAAGTCACGGAGTGTGAATAGTCAGAG	
SC-5	30mer	
	ACACGAGACTGTTTGTTTGAAGACCCCTTG	
SC-6	25mer	
	TGGAATCTCAACAAATATACAGGTG	
SN-3RV	30mer	
	GGGTCATGGCCAACGTGGACGTATAAGCAC	
SN-4RV	30mer	
	GATGATCACTGGCGCGGTTTTCTCCTCGAG	

Example 10 (Acquisition of Raffinose Synthetase Gene from Japanese Artichoke cDNA)

In the same manner as described in Example 5, cDNA was obtained from leaves of Japanese artichoke (*Stachys sieboldii*). Using this cDNA as a template and primers designed from the amino acid sequence of SEQ ID NO:1, i.e., primers having nucleotide sequences shown in list 10 below, a DNA fragment was amplified by PCR under the conditions described in Example 6. The DNA fragment thus amplified by PCR was cloned with a TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and nucleotide sequence analysis with a 373S DNA sequencer of ABI. As a result, the nucleotide sequence of SEQ ID:6 was determined.

Based on the nucleotide sequence thus obtained, synthesized DNA primers are prepared, and in the same manner as described in Example 9, the nucleotide sequences in both terminal regions of the gene are analyzed with Marathon Kit of Clontech.

(List 10)

1-F primer	35mer	
	CGATTIAAIGTITGGTGGACIACICAITGGGTIGG	
4-RV primer	37mer	
	GGCCAGCIATIACICCITTICCITTIAAITGITTITT	
2-F primer	44mer	
	CGAATIATIGAIAAITTIGGITGGTGIACITGGGAIGCITTITA	
6-RV primer	32mer	
	GGCCACATITTIACIA(AG)ICCIATIGGIGCIAA	

Example 11 (Acquisition of Raffinose Synthetase Gene from Corn cDNA)

In the same manner as described in Example 5, cDNA was obtained from leaves of corn (*Zea mays* L.) Pioneer 3358. Using this cDNA as a template and primers designed from the amino acid sequence of SEQ ID NO:1, i.e., primers having nucleotide sequences shown in list 11 below, a DNA fragment was amplified by PCR under the conditions described in Example 6. The DNA fragment thus amplified by PCR was cloned with a TA cloning kit (Invitrogen), followed by sequence reaction using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit of Perkin-Elmer and nucleotide sequence analysis with a 373S DNA sequencer of ABI. Based on this sequence, primers having nucleotide sequences shown in list 12 below were synthesized. In the same manner as described in Example 5, mRNA obtained from leaves of corn (*Zea mays* L.) Pioneer 3358 was linked to an adaptor contained in the Marathon Kit of Clontech with ligase. This operation was made according to the protocol attached. Using the adaptor-ligated cDNA thus prepared, polymerase chain reaction was effected in the same manner as described above with the primers shown in list 12 below. As a result, the nucleotide sequence of SEQ ID NO:8 was determined.

Based on the nucleotide sequence thus obtained, synthesized DNA primers are prepared, and in the same manner as described in Example 9, the nucleotide sequence in the 5'-terminal region of the gene is analyzed with Marathon Kit of Clontech.

(List 11)

5-F primer 41mer
CGATGGATGGGIAAITTIATICAICCIGAITGGGAIATGTT
6-RV primer 32mer
GGCCACATITTIACIA(AG)ICCIATIGGIGCIAA

(List 12)

M-10 primer 25mer
GACGTCGAGTGGAAGAGCGGCAAGG
M-11 primer 25mer
CACCTACGAGCTCTTCGTCGTTGCC

Example 12 (Construction of Expression Vectors in Plant for Chimera Gene, 35S-Broad Bean Raffinose Synthetase Gene)

The plasmid pBluescriptKS-RS having the broad bean raffinose synthetase gene obtained in Example 7 was digested with the restriction endonucleases *Bam* HI and *Sac* I. Using Ligation Kit (Takara), the DNA fragment thus digested was cloned in the binary vector pBI121 (Clontech) previously digested with *Bam* HI and *Sac* I. The vector thus obtained was designated pBI121-RS.

For an antisense experiment, plasmid pBI121 (Clontech) previously digested with *Bam* HI and *Sac* I was ligated to linkers shown in list 13 below to give pBI121(-). This pBI121(-) was used to prepare pBI121(-)-RS in the same manner as described for the preparation of pBI121-RS above.

A similar vector was prepared with pBI221. The plasmid pBluescriptKS-RS obtained in Example 7 was digested with the restriction endonucleases *Bam* HI and *Sac* I. Using Ligation Kit (Takara), the DNA fragment thus digested was cloned in the vector pBI221 (Clontech) previously digested with *Bam* HI and *Sac* I. The vector thus obtained was designated pBI221-RS.

For an antisense experiment, plasmid pBI221 (Clontech) previously digested with *Bam* HI and *Sac* I was ligated to linkers shown in list 13 below to give pBI221(-). This pBI221(-) was used to prepare pBI221(-)-RS in the same manner as described for the preparation of pBI221-RS above.

The construction of these expression vectors is shown in Figures 2 and 3.

(List 13)

BamSac-(+) linker 25mer
GATCGAGCTCGTGTCGGATCCAGCT
BamSac-(-) linker 17mer
GGATCCGACACGAGCTC

Example 13 (Transformation of Mustard with Broad Bean Raffinose Synthetase Gene)

The vectors pBI121-RS and pBI121(-)-RS prepared in Example 12 were used for the transformation of mustard (*Brassica juncia*) by the Agrobacterium infection method.

Agrobacterium tumefaciens (strain C58C1, rifampicin resistant) previously made into a competent state by calcium chloride treatment was transformed independently with two plasmids pBI121-RS and pBI121(-)-RS prepared in Example 12. Selection for transformants was carried out on LB medium containing 50 μg/ml rifampicin and 25 μg/ml kanamycin by utilizing the character of kanamycin resistance conferred by the kanamycin resistance gene (neomycin phosphotransferase, NPTII) of the introduced plasmids.

The transformant Agrobacterium obtained (*Agrobacterium tumefaciens* strain C58, rifampicin resistant) was cultivated on LB medium containing 50 μg/ml rifampicin and 25 μg/ml kanamycin at 28°C for a whole day and night, and the culture was used for the transformation of mustard by the method described below.

The seeds of mustard were aseptically sowed on 1/2 MS medium, 2% sucrose, 0.7% agar. After one week, cotyledons and petioles of sprouting plants were cut out with a scalpel, and transferred to MS medium, 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D, 3.3 μM $AgNO_3$, followed by precultivation for 1 day. The precultivated cotyledons and petioles were transferred in a 1000-fold dilution of the Agrobacterium culture to cause infection for 5 minutes. The infected cotyledons and petioles were transferred again to the same medium as used in the precultivation, and cultivated for 3 to 4 days. The cultivated cotyledons and petioles were transferred to MS medium, 3% sucrose, 4.5 μM BA,

0.05 μM 2.4-D, 3.3 μM $AgNO_3$, 500 mg/l cefotaxim, and sterilized with shaking for 1 day. The sterilized cotyledons and petioles were transferred to MS medium, 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D, 3.3 μM $AgNO_3$, 100 mg/l cefotaxim, 20 mg/l kanamycin, and cultivated for 3 to 4 weeks. The cotyledons and petioles were transferred to MS medium, 3% sucrose, 0.7% agar, 4.5 μM BA, 0.05 μM 2.4-D, 100 mg/l cefotaxim, 20 mg/l kanamycin, and cultivated. The cultivation on this medium was continued with subculturing at intervals of 3 to 4 weeks. When the regeneration of shoots began to occur, these shoots were subcultured on MS medium, 3% sucrose, 0.7% agar, 20 mg/l kanamycin, and cultivated for 3 to 4 weeks. The rooting plants were transferred to vermiculite : peat moss = 1 : 1, and conditioned at 21° to 22°C in a cycle of day/night = 12 hours : 12 hours. With the progress of plant body growth, the plants were suitably grown with cultivation soil. From leaves of the regenerated plants, genomic DNA was extracted according to the method described above, and the gene insertion into the plant genome was confirmed by PCR using the primers shown in list 14 below.

(List 14)

35S	30mer	
	TTCCAGTATGGACGATTCAAGGCTTGCTTC	
NOS	25mer	
	ATGTATAATTGCGGGACTCTAATCA	
RS-F	30mer	
	AAGAGTGTATCTGAATTTTCACGCGCGGTG	
RS-RV	33mer	
	ACCTTCCCATACACCTTTTGGATGAACCTTCAA	

Example 14 (Transformation of Soybean Somatic Embryo with Broad Bean Raffinose Synthetase Gene)

Cultured cells of soybean "Fayette" somatic embryos (400 to 500 mg FW) were arranged in one layer within a circle having a diameter of 20 mm on the central part of a 6 cm agar plate. Two plasmids pBI221-RS and pBI221(-)-RS having chimera genes prepared from the broad bean raffinose synthetase gene and 35S promoter in Example 12 were introduced into the soybean somatic embryos according to the disclosure of the Japanese Patent Application No. 3-291501/1991. That is, these plasmids were mixed with the β-glucuronidase (GUS)/hygromycin-resistant gene (HPT) coexpression vector pSUM-GH:NotI for selection described in Soshiki Baiyo, 20, 323-327 (1994). These mixed plasmids were used for the gene introduction into the soybean somatic embryos with a particle gun (800 mg/coating gold particles 200 μg/shot; projectile stopper-sample distance, 100 mm). After the introduction, gyratory cultures were grown in the MS modified growth liquid medium (Sigma) containing 25 to 50 μg/ml hygromycin under illumination at 25°C for 16 hours, and transformed somatic embryos were selected.

For the hygromycin-resistant soybean somatic embryos having yellowish green color and growth ability, which were selected after about 3 months, polymerization chain reaction is effected with primers shown in list 14 above to determine whether the broad bean raffinose synthetase gene region is amplified or not. This confirms that the broad bean raffinose synthetase gene is inserted into the soybean genome.

Furthermore, the somatic embryos obtained are used for the regeneration of plants to give transformant soybean with the broad bean raffinose synthetase gene.

Medium Composition

LB and MS media used in the above Examples have the following respective compositions.

(LB medium)	
Bacto-tryptone	10 g
Bacto-yeast extract	5 g
NaCl	10 g / 1 liter H_2O (pH 7.0)

(MS medium)	
KNO_3	2022 mg/l
NH_4NO_3	1650 mg/l
NH_4Cl	2140 mg/l
KH_2PO_4	170 mg/l
$MgSO_4 \cdot 7H_2O$	370 mg/l
$CaCl_2 \cdot 2H_2O$	440 mg/l
$MnSO_4 \cdot 4H_2O$	22.3 mg/l
$ZnSO_4 \cdot 7H_2O$	8.6 mg/l
$CuSO_4 \cdot 5H_2O$	0.025 mg/l
KI	0.83 mg/l
$CoCl_2 \cdot 6H_2O$	0.025 mg/l
H_3BO_3	6.2 mg/l
NaMoO4 · $2H_2O$	0.25 mg/l
$FeSO_4 \cdot 7H_2O$	27.8 mg/l
Na_2EDTA	37.3 mg/l
Nicotinic acid	0.5 mg/l
Thiamine HCl	1 mg/l
Pyridoxine HCl	0.5 mg/l
Inositol	100 mg/l
Glycine	2 mg/l

Brief Description of the Sequences

1. SEQ ID NO:1:

The sequence of SEQ ID NO:1 shows an amino acid sequence of a raffinose synthetase protein encoded in the raffinose synthetase gene obtained from broad bean.

2. SEQ ID NO:2:

The sequence of SEQ ID NO:2 shows a cDNA nucleotide sequence of the raffinose synthetase gene obtained from broad bean.

3. SEQ ID NO:3:

The sequence of SEQ ID NO:3 shows an amino acid sequence of a raffinose synthetase protein encoded in the raffinose synthetase gene obtained from soybean.

4. SEQ ID NO:4:

The sequence of SEQ ID NO:4 shows a cDNA nucleotide sequence of the raffinose synthetase gene obtained from soybean.

5. SEQ ID NO:5:

The sequenc of SEQ ID NO:5 shows an amino acid sequenc of a raffinose synthetase protein encoded in the raffinose synthetase gene obtained from Japanese artichoke.

6. SEQ ID NO:6:

The sequence of SEQ ID NO:6 shows a cDNA nucleotide sequence of the raffinose synthetase gene obtained from Japanese artichoke.

7. SEQ ID NO:7:

The sequence of SEQ ID NO:7 shows an amino acid sequence of a raffinose synthetase protein encoded in the raffinose synthetase gene obtained from corn.

8. SEQ ID NO:8:

The sequence of SEQ ID NO:8 shows a cDNA nucleotide sequence of the raffinose synthetase gene obtained from corn.

9. List 1:

The nucleotide sequences shown in list 1 are of the typical primers used in the amplification of a cDNA fragment of a raffinose synthetase gene. All of these sequences are based on the nucleotide sequence in the non-coding region of the gene. Primer 1 is a sense primer corresponding to the 5'-terminus of a cDNA fragment of the broad bean-derived raffinose synthetase gene. Primers 2 and 3 are antisense primers corresponding to the 3'-terminus of the cDNA fragment of the broad bean-derived raffinose synthetase gene. Primer 4 is a sense primer corresponding to the 5'-terminus of a cDNA fragment of the soybean-derived raffinose synthetase gene. Primers 5 and 6 are antisense primers corresponding to the 3'-terminus of the cDNA fragment of the soybean-derived raffinose synthetase gene. Depending upon the purpose, recognition sequences for suitable restriction endonucleases can be added to the 5'-termini of these nucleotide sequences in an appropriate manner.

10. List 2:

The nucleotide sequences shown in list 2 are of the typical primers used in the amplification of an open reading frame coding for the amino acid sequence of a raffinose synthetase protein in the cDNA sequence of a raffinose synthetase gene. Primers 1 and 2 are sense primers corresponding to the N-terminus of the broad bean-derived raffinose synthetase protein. Primers 3 and 4 are antisense primers corresponding to the C-terminus of the broad bean-derived raffinose synthetase protein. Primers 5 and 6 are sense primers corresponding to the N-terminus of the soybean-derived raffinose synthetase protein. Primers 7 and 8 are antisense primers corresponding to the C-terminus of the soybean-derived raffinose synthetase protein. Depending upon the purpose, recognition sequences for suitable restriction endonucleases can be added to the 5'-termini of these sequences in an appropriate manner.

11. List 3:

The amino acid sequences shown in list 3 are partial amino acid sequences of a raffinose synthetase protein.

#1 is equivalent to the partial amino acid sequence extending from amino acid 110 to amino acid 129 in the amino acid sequence of SEQ ID NO:1.
#2 is equivalent to the partial amino acid sequence extending from amino acid 234 to amino acid 247 in the amino acid sequence of SEQ ID NO:1.
#3 is equivalent to the partial amino acid sequence extending from amino acid 265 to amino acid 279 in the amino acid sequence of SEQ ID NO:1.
#4 is equivalent to the partial amino acid sequence extending from amino acid 296 to amino acid 312 in the amino acid sequence of SEQ ID NO:1.
#5 is equivalent to the partial amino acid sequence extending from amino acid 346 to amino acid 361 in the amino acid sequenc of SEQ ID NO:1.
#6 is equivalent to the partial amino acid sequence extending from amino acid 383 to amino acid 402 in the amino

acid sequence of SEQ ID NO:1.

#7 is equivalent to the partial amino acid sequence extending from amino acid 411 to amino acid 433 in the amino acid sequence of SEQ ID NO:1.

#8 is equivalent to the partial amino acid sequence extending from amino acid 440 to amino acid 453 in the amino acid sequence of SEQ ID NO:1.

#9 is equivalent to the partial amino acid sequence extending from amino acid 457 to amino acid 468 in the amino acid sequence of SEQ ID NO:1.

#10 is equivalent to the partial amino acid sequence extending from amino acid 471 to amino acid 516 in the amino acid sequence of SEQ ID NO:1.

#11 is equivalent to the partial amino acid sequence extending from amino acid 517 to amino acid 559 in the amino acid sequence of SEQ ID NO:1.

#12 is equivalent to the partial amino acid sequence extending from amino acid 574 to amino acid 582 in the amino acid sequence of SEQ ID NO:1.

#13 is equivalent to the partial amino acid sequence extending from amino acid 586 to amino acid 609 in the amino acid sequence of SEQ ID NO:1.

#14 is equivalent to the partial amino acid sequence extending from amino acid 615 to amino acid 627 in the amino acid sequence of SEQ ID NO:1.

#15 is equivalent to the partial amino acid sequence extending from amino acid 716 to amino acid 724 in the amino acid sequence of SEQ ID NO:1.

12. List 4:

The nucleotide sequences shown in list 4 are of the typical primers synthesized on some of the amino acid sequences shown in list 3. The symbol "F" as used after the primer number means that the primer referred to by this symbol has a sense sequence. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence. Primer 1 corresponds to the partial amino acid sequence extending from amino acid 119 to amino acid 129 in the amino acid sequence of SEQ ID NO:1. Primer 2 corresponds to the partial amino acid sequence extending from amino acid 234 to amino acid 247 in the amino acid sequence of SEQ ID NO:1. Primer 3 corresponds to the partial amino acid sequence extending from amino acid 265 to amino acid 279 in the amino acid sequence of SEQ ID NO:1. Primer 4 corresponds to the partial amino acid sequence extending from amino acid 458 to amino acid 468 in the amino acid sequence of SEQ ID NO:1. Primer 5 corresponds to the partial amino acid sequence extending from amino acid 522 to amino acid 534 in the amino acid sequence of SEQ ID NO:1. Primer 6 corresponds to the partial amino acid sequence extending from amino acid 716 to amino acid 724 in the amino acid s quence of SEQ ID NO:1.

13. List 5:

The nucleotide sequences shown in list 5 are of the typical primers synthesized on the partial amino acid sequences of the purified broad bean raffinose synthetase protein. The bases shown in parentheses mean that a mixture of those bases was used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

14. List 6:

The nucleotide sequences shown in list 6 are of the typical primers used in the analysis of both terminal regions of a cDNA nucleotide sequence of the broad bean raffinose synthetase gene by the RACE method. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

15. List 7:

The nucleotide sequences shown in list 7 are of the typical primers used in the cloning of the broad bean raffinose synthetase gene. RS-N corresponds to the N-terminus of the open reading frame and contains two recognition sites for the restriction endonucleases *Bam* HI and *Nco* I on the 5'-terminal side. RS-C is an antisense primer corresponding to the C-terminus of the open reading frame and contains a recognition site for the restriction endonuclease *Xba* I on the 5'-terminal side.

16. List 8:

The nucleotide sequences shown in list 8 are of the typical primers used in the cloning of a soybean raffinose synthetase gene fragment. The base represented by the symbol "I" was inosine used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

17. List 9:

The nucleotide sequences shown in list 9 are of the typical primers used in the analysis of the cDNA nucleotide sequence of a soybean raffinose synthetase gene fragment. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

The analysis of nucleotide sequences was carried out by polymerase chain reaction using SN-1 and SC-3RV. SC-5 and SC-6 were used in the analysis of a nucleotide sequence in the 3'-terminal region, and SN-3RV and SN-4RV were used in the analysis of a nucleotide sequence in the 5'-terminal region.

18. List 10:

The nucleotide sequences shown in list 10 are of the typical primers used in the analysis of the cDNA nucleotide sequence of a Japanese artichoke raffinose synthetase gene fragment. The base represented by the symbol "I" was inosine used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

19. List 11:

The nucleotide sequences shown in list 11 are of the typical primers used in the analysis of the cDNA nucleotide sequence of a corn raffinose synthetase gene fragment. The base represented by the symbol "I" was inosine used in the synthesis. The symbol "RV" as used after the primer number means that the primer referred to by this symbol has an antisense sequence.

20. List 12:

The nucleotide sequences shown in list 12 are of the typical primers used in the analysis of the cDNA nucleotide sequence of a corn raffinose synthetase gene fragment. M-10 and M-11 were used in the analysis of a nucleotide sequence in the 3'-terminal region.

21. List 13:

The nucleotide sequences shown in list 13 are of the typical adopters used in the construction of vectors for antisense experiments. These synthetic DNA fragments takes a double-stranded form when mixed together because they are complementary strands. This double-stranded DNA fragment has cohesive ends of cleavage sites for the restriction endonucleases *Bam* HI and *Sac* I on both termini, and contains the restriction sites for the restriction endonucleases *Bam* HI and *Sac* I in the double-stranded region.

22. List 14:

The nucleotide sequences shown in list 14 are of the typical primers used in the PCR experiment to confirm the gene introduction into the genome of a recombinant plant. 35S is a primer toward the downstream region at the 35S promoter site, and NOS is a primer toward the upstream region at the NOS terminator site. RS-F is a sense primer of the broad bean raffinose synthetase gene, and RS-RV is an antisense primer of the broad bean raffinose synthetase gene.

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME:Sumitomo Chemical Co., Ltd.
(B) STREET: 5-33, Kitahama 4-chome, Chuo-ku
(C) CITY:Osaka-shi, Osaka-fu
(E) COUNTRY: Japan
(F) POSTAL CODE (ZIP): none

(ii) TITLE OF INVENTION: Raffinose synthetase genes and the use thereof

(iii) NUMBER OF SEQUENCES: 8

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PS-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0,Version #1.25(EPO)

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 97 12 2417.5
(B) FILING DATE: 18-DEC-1997

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 338673/1996
(B) FILING DATE: 18-DEC-1996

(1) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 799 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: broad bean (Vicia faba)
(B) STRAIN: Nintoku Issun
(F) TISSUE TYPE: seeds

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Ala Pro Pro Ser Ile Thr Lys Thr Ala Thr Leu Gln Asp Val Ile
 1               5                  10                  15
Ser Thr Ile Asp Ile Gly Asn Gly Asn Ser Pro Leu Phe Ser Ile Thr
            20                  25                  30
Leu Asp Gln Ser Arg Asp Phe Leu Ala Asn Gly His Pro Phe Leu Thr
        35                  40                  45
Gln Val Pro Pro Asn Ile Thr Thr Thr Thr Thr Thr Thr Ala Ser Ser
    50                  55                  60
Phe Leu Asn Leu Lys Ser Asn Lys Asp Thr Ile Pro Asn Asn Asn Asn
65                  70                  75                  80
```

```
Thr Met Leu Leu Gln Gln Gly Cys Phe Val Gly Phe Asn Ser Thr Glu
                  85                  90                  95
Pro Lys Ser His His Val Val Pro Leu Gly Lys Leu Lys Gly Ile Lys
             100                 105                 110
Phe Met Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp Val
         115                 120                 125
Gly Thr Asn Gly Gln Glu Leu Gln His Glu Thr Gln Met Leu Ile Leu
     130                 135                 140
Asp Lys Asn Asp Ser Leu Gly Arg Pro Tyr Val Leu Leu Leu Pro Ile
145                 150                 155                 160
Leu Glu Asn Thr Phe Arg Thr Ser Leu Gln Pro Gly Leu Asn Asp His
                 165                 170                 175
Ile Gly Met Ser Val Glu Ser Gly Ser Thr His Val Thr Gly Ser Ser
             180                 185                 190
Phe Lys Ala Cys Leu Tyr Ile His Leu Ser Asn Asp Pro Tyr Ser Ile
         195                 200                 205
Leu Lys Glu Ala Val Lys Val Ile Gln Thr Gln Leu Gly Thr Phe Lys
     210                 215                 220
Thr Leu Glu Glu Lys Thr Ala Pro Ser Ile Ile Asp Lys Phe Gly Trp
225                 230                 235                 240
Cys Thr Trp Asp Ala Phe Tyr Leu Lys Val His Pro Lys Gly Val Trp
                 245                 250                 255
Glu Gly Val Lys Ser Leu Thr Asp Gly Gly Cys Pro Pro Gly Phe Val
             260                 265                 270
Ile Ile Asp Asp Gly Trp Gln Ser Ile Cys His Asp Asp Asp Asp Glu
         275                 280                 285
Asp Asp Ser Gly Met Asn Arg Thr Ser Ala Gly Glu Gln Met Pro Cys
     290                 295                 300
Arg Leu Val Lys Tyr Glu Glu Asn Ser Lys Phe Arg Glu Tyr Glu Asn
305                 310                 315
Pro Glu Asn Gly Gly Lys Lys Gly Leu Gly Gly Phe Val Arg Asp Leu
320                 325                 330                 335
Lys Glu Glu Phe Gly Ser Val Glu Ser Val Tyr Val Trp His Ala Leu
                 340                 345                 350
Cys Gly Tyr Trp Gly Gly Val Arg Pro Gly Val His Gly Met Pro Lys
             355                 360                 365
Ala Arg Val Val Val Pro Lys Val Ser Gln Gly Leu Lys Met Thr Met
     370                 375                 380
Glu Asp Leu Ala Val Asp Lys Ile Val Glu Asn Gly Val Gly Leu Val
385                 390                 395                 400
Pro Pro Asp Phe Ala His Glu Met Phe Asp Gly Leu His Ser His Leu
                 405                 410                 415
Glu Ser Ala Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu
             420                 425                 430
Glu Leu Leu Ser Glu Glu Tyr Gly Gly Arg Val Glu Leu Ala Arg Ala
         435                 440                 445
Tyr Tyr Lys Ala Leu Thr Ser Ser Val Lys Lys His Phe Lys Gly Asn
     450                 455                 460
Gly Val Ile Ala Ser Met Glu His Cys Asn Asp Phe Phe Leu Leu Gly
465                 470                 475                 480
Thr Glu Ala Ile Ser Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Ser
                 485                 490                 495
Asp Pro Ser Gly Asp Pro Asn Gly Thr Tyr Trp Leu Gln Gly Cys His
             500                 505                 510
Met Val His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln
         515                 520                 525
Pro Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His
     530                 535                 540
```

```
Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Cys
545                 550                 555                 560
Val Gly Asn His Asn Phe Lys Leu Leu Lys Ser Leu Val Leu Pro Asp
                    565                 570                 575
Gly Ser Ile Leu Arg Cys Gln His Tyr Ala Leu Pro Thr Arg Asp Cys
            580                 585                 590
Leu Phe Glu Asp Pro Leu His Asn Gly Lys Thr Met Leu Lys Ile Trp
        595                 600                 605
Asn Leu Asn Lys Tyr Thr Gly Val Leu Gly Leu Phe Asn Cys Gln Gly
    610                 615                 620
Gly Gly Trp Cys Pro Glu Ala Arg Arg Asn Lys Ser Val Ser Glu Phe
625                 630                 635                 640
Ser Arg Ala Val Thr Cys Tyr Ala Ser Pro Glu Asp Ile Glu Trp Cys
                645                 650                 655
Asn Gly Lys Thr Pro Met Ser Thr Lys Gly Val Asp Phe Phe Ala Val
            660                 665                 670
Tyr Phe Phe Lys Glu Lys Lys Leu Arg Leu Met Lys Cys Ser Asp Arg
        675                 680                 685
Leu Lys Val Ser Leu Glu Pro Phe Ser Phe Glu Leu Met Thr Val Ser
    690                 695                 700
Pro Val Lys Val Phe Ser Lys Arg Phe Ile Gln Phe Ala Pro Ile Gly
705                 710                 715                 720
Leu Val Asn Met Leu Asn Ser Gly Gly Ala Ile Gln Ser Leu Glu Phe
                725                 730                 735
Asp Asp Asn Ala Ser Leu Val Lys Ile Gly Val Arg Gly Cys Gly Glu
            740                 745                 750
Met Ser Val Phe Ala Ser Glu Lys Pro Val Cys Cys Lys Ile Asp Gly
        755                 760                 765
Val Lys Val Lys Phe Leu Tyr Glu Asp Lys Met Ala Arg Val Gln Ile
    770                 775                 780
Leu Trp Pro Ser Ser Ser Thr Leu Ser Leu Val Gln Phe Leu Phe Stop
785                 790                 795                 800
```

(1) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2746 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: peptide
(B) LOCATION: 101..2500
(C) IDENTIFICATION METHOD: by experiment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
AATTTTCAAG CATAGCCAAG TTAACCACCT TAGAAACATT CCTACAAGCT ACTTATCCCT    60
GTCAATAAGC TACTAAGCTA CCAGAGTCTC ATCAATCACC ATG GCA CCA CCA AGC     115
                                             Met Ala Pro Pro Ser
                                                                5
ATA ACC AAA ACT GCA ACC CTC CAA GAC GTA ATA AGC ACC ATC GAT ATT     163
Ile Thr Lys Thr Ala Thr Leu Gln Asp Val Ile Ser Thr Ile Asp Ile
                10                  15                  20
```

```
GGT AAT GGT AAC TCA CCC TTA TTC TCC ATA ACC TTA GAC CAA TCA CGT      211
Gly Asn Gly Asn Ser Pro Leu Phe Ser Ile Thr Leu Asp Gln Ser Arg
                25                  30                  35
GAC TTC CTT GCA AAT GGC CAC CCT TTC CTC ACC CAA GTC CCA CCT AAC      259
Asp Phe Leu Ala Asn Gly His Pro Phe Leu Thr Gln Val Pro Pro Asn
        40                  45                  50
ATA ACA ACA ACA ACA ACA ACC ACT GCT TCC TCT TTT CTC AAT CTC AAA      307
Ile Thr Thr Thr Thr Thr Thr Thr Ala Ser Ser Phe Leu Asn Leu Lys
    55                  60                  65
TCC AAC AAA GAT ACC ATT CCC AAC AAC AAC AAC ACC ATG TTG TTG CAA      355
Ser Asn Lys Asp Thr Ile Pro Asn Asn Asn Asn Thr Met Leu Leu Gln
70                  75                  80                  85
CAA GGT TGT TTC GTT GGT TTC AAC TCC ACC GAA CCC AAA AGC CAC CAC      403
Gln Gly Cys Phe Val Gly Phe Asn Ser Thr Glu Pro Lys Ser His His
                90                  95                  100
GTA GTT CCA CTC GGC AAA CTA AAA GGA ATC AAA TTC ATG AGC ATA TTC      451
Val Val Pro Leu Gly Lys Leu Lys Gly Ile Lys Phe Met Ser Ile Phe
            105                 110                 115
CGG TTC AAA GTT TGG TGG ACA ACT CAC TGG GTC GGA ACA AAT GGA CAG      499
Arg Phe Lys Val Trp Trp Thr Thr His Trp Val Gly Thr Asn Gly Gln
        120                 125                 130
GAA CTA CAA CAC GAA ACA CAA ATG TTA ATC CTG GAC AAA AAC GAC TCC      547
Glu Leu Gln His Glu Thr Gln Met Leu Ile Leu Asp Lys Asn Asp Ser
    135                 140                 145
CTC GGA CGA CCC TAT GTC TTA CTC CTC CCA ATC CTA GAA AAC ACC TTC      595
Leu Gly Arg Pro Tyr Val Leu Leu Leu Pro Ile Leu Glu Asn Thr Phe
150                 155                 160                 165
CGA ACC TCA CTC CAA CCC GGT CTC AAC GAT CAC ATA GGC ATG TCC GTC      643
Arg Thr Ser Leu Gln Pro Gly Leu Asn Asp His Ile Gly Met Ser Val
                170                 175                 180
GAA AGC GGT TCA ACA CAT GTC ACC GGG TCA AGC TTC AAA GCA TGT CTT      691
Glu Ser Gly Ser Thr His Val Thr Gly Ser Ser Phe Lys Ala Cys Leu
            185                 190                 195
TAC ATC CAT CTC AGT AAC GAC CCA TAC AGT ATA CTA AAA GAA GCA GTT      739
Tyr Ile His Leu Ser Asn Asp Pro Tyr Ser Ile Leu Lys Glu Ala Val
        200                 205                 210
AAA GTA ATC CAA ACT CAG TTA GGA ACA TTC AAG ACT CTT GAA GAA AAA      787
Lys Val Ile Gln Thr Gln Leu Gly Thr Phe Lys Thr Leu Glu Glu Lys
    215                 220                 225
ACA GCA CCT AGT ATT ATA GAC AAA TTC GGT TGG TGC ACG TGG GAT GCT      835
Thr Ala Pro Ser Ile Ile Asp Lys Phe Gly Trp Cys Thr Trp Asp Ala
230                 235                 240                 245
TTT TAC TTG AAG GTT CAT CCA AAA GGT GTA TGG GAA GGT GTA AAG TCT      883
Phe Tyr Leu Lys Val His Pro Lys Gly Val Trp Glu Gly Val Lys Ser
                250                 255                 260
CTC ACA GAT GGT GGT TGT CCT CCC GGT TTC GTC ATA ATC GAC GAC GGT      931
Leu Thr Asp Gly Gly Cys Pro Pro Gly Phe Val Ile Ile Asp Asp Gly
            265                 270                 275
TGG CAA TCC ATT TGT CAT GAC GAT GAC GAT GAA GAT GAT TCA GGA ATG      979
Trp Gln Ser Ile Cys His Asp Asp Asp Asp Glu Asp Asp Ser Gly Met
        280                 285                 290
AAC CGA ACC TCA GCC GGG GAA CAA ATG CCA TGC AGA CTT GTA AAA TAC     1027
Asn Arg Thr Ser Ala Gly Glu Gln Met Pro Cys Arg Leu Val Lys Tyr
    295                 300                 305
GAA GAG AAT TCT AAG TTT AGA GAA TAT GAG AAT CCT GAA AAT GGA GGG     1075
Glu Glu Asn Ser Lys Phe Arg Glu Tyr Glu Asn Pro Glu Asn Gly Gly
310                 315                 320                 325
```

```
AAG AAA GGT TTG GGT GGT TTT GTG AGG GAT TTG AAG GAA GAG TTT GGG    1123
Lys Lys Gly Leu Gly Gly Phe Val Arg Asp Leu Lys Glu Glu Phe Gly
                330                 335                 340
AGT GTG GAG AGT GTT TAT GTT TGG CAT GCG CTT TGT GGG TAT TGG GGC    1171
Ser Val Glu Ser Val Tyr Val Trp His Ala Leu Cys Gly Tyr Trp Gly
            345                 350                 355
GGG GTT AGG CCT GGA GTG CAT GGG ATG CCG AAA GCT AGG GTT GTT GTT    1219
Gly Val Arg Pro Gly Val His Gly Met Pro Lys Ala Arg Val Val Val
        360                 365                 370
CCG AAG GTG TCT CAG GGG TTG AAG ATG ACG ATG GAG GAT TTG GCG GTG    1267
Pro Lys Val Ser Gln Gly Leu Lys Met Thr Met Glu Asp Leu Ala Val
    375                 380                 385
GAT AAG ATT GTT GAG AAC GGT GTG GGG CTA GTG CCG CCA GAT TTT GCA    1315
Asp Lys Ile Val Glu Asn Gly Val Gly Leu Val Pro Pro Asp Phe Ala
390                 395                 400                 405
CAT GAG ATG TTT GAT GGG CTT CAC TCT CAT TTG GAG TCG GCG GGA ATT    1363
His Glu Met Phe Asp Gly Leu His Ser His Leu Glu Ser Ala Gly Ile
                410                 415                 420
GAC GGT GTT AAA GTT GAT GTT ATC CAT CTG CTT GAG TTA CTA TCA GAG    1411
Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu Leu Leu Ser Glu
            425                 430                 435
GAA TAT GGT GGA CGA GTT GAG CTA GCA AGA GCT TAT TAC AAA GCA CTA    1459
Glu Tyr Gly Gly Arg Val Glu Leu Ala Arg Ala Tyr Tyr Lys Ala Leu
        440                 445                 450
ACC TCA TCA GTG AAG AAA CAT TTC AAA GGC AAT GGT GTA ATT GCT AGC    1507
Thr Ser Ser Val Lys Lys His Phe Lys Gly Asn Gly Val Ile Ala Ser
    455                 460                 465
ATG GAG CAT TGC AAC GAC TTC TTT CTC CTC GGC ACC GAA GCC ATA TCC    1555
Met Glu His Cys Asn Asp Phe Phe Leu Leu Gly Thr Glu Ala Ile Ser
470                 475                 480                 485
CTC GGC CGC GTC GGA GAT GAT TTT TGG TGC TCT GAT CCA TCT GGT GAT    1603
Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Ser Asp Pro Ser Gly Asp
                490                 495                 500
CCA AAT GGT ACA TAT TGG CTC CAA GGT TGT CAC ATG GTA CAT TGT GCC    1651
Pro Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val His Cys Ala
            505                 510                 515
TAC AAC AGT TTA TGG ATG GGA AAT TTC ATT CAG CCA GAT TGG GAC ATG    1699
Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro Asp Trp Asp Met
        520                 525                 530
TTT CAG TCC ACT CAT CCT TGT GCT GAA TTT CAT GCC GCC TCA CGA GCC    1747
Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala Ser Arg Ala
    535                 540                 545
ATA TCC GGC GGA CCA ATT TAT GTT AGT GAT TGT GTT GGT AAT CAC AAT    1795
Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Cys Val Gly Asn His Asn
550                 555                 560                 565
TTC AAG TTG CTC AAA TCT CTT GTT TTG CCC GAT GGT TCT ATC TTG CGT    1843
Phe Lys Leu Leu Lys Ser Leu Val Leu Pro Asp Gly Ser Ile Leu Arg
                570                 575                 580
TGT CAA CAT TAC GCA CTC CCT ACA AGA GAT TGC TTG TTT GAA GAC CCT    1891
Cys Gln His Tyr Ala Leu Pro Thr Arg Asp Cys Leu Phe Glu Asp Pro
            585                 590                 595
TTG CAT AAT GGC AAA ACA ATG CTG AAA ATT TGG AAT CTC AAC AAA TAT    1939
Leu His Asn Gly Lys Thr Met Leu Lys Ile Trp Asn Leu Asn Lys Tyr
        600                 605                 610
ACA GGT GTT TTG GGT CTT TTC AAC TGC CAA GGT GGT GGG TGG TGT CCT    1987
Thr Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly Gly Trp Cys Pro
    615                 620                 625
```

```
GAG GCA CGG CGA AAC AAG AGT GTA TCT GAA TTT TCA CGC GCG GTG ACA    2035
Glu Ala Arg Arg Asn Lys Ser Val Ser Glu Phe Ser Arg Ala Val Thr
630                 635                 640                 645
TGT TAT GCA AGT CCC GAA GAC ATT GAA TGG TGC AAT GGG AAA ACT CCA    2083
Cys Tyr Ala Ser Pro Glu Asp Ile Glu Trp Cys Asn Gly Lys Thr Pro
                650                 655                 660
ATG AGC ACC AAA GGT GTG GAT TTT TTT GCT GTG TAT TTT TTC AAG GAG    2131
Met Ser Thr Lys Gly Val Asp Phe Phe Ala Val Tyr Phe Phe Lys Glu
            665                 670                 675
AAG AAA TTG AGG CTC ATG AAG TGT TCT GAT AGA TTG AAA GTT TCG CTT    2179
Lys Lys Leu Arg Leu Met Lys Cys Ser Asp Arg Leu Lys Val Ser Leu
        680                 685                 690
GAG CCA TTT AGT TTT GAG CTA ATG ACA GTG TCT CCA GTG AAA GTG TTT    2227
Glu Pro Phe Ser Phe Glu Leu Met Thr Val Ser Pro Val Lys Val Phe
    695                 700                 705
TCG AAA AGG TTT ATA CAG TTT GCA CCG ATT GGG TTA GTG AAC ATG CTG    2275
Ser Lys Arg Phe Ile Gln Phe Ala Pro Ile Gly Leu Val Asn Met Leu
710                 715                 720                 725
AAC TCT GGT GGT GCG ATT CAG TCT CTG GAG TTT GAT GAT AAT GCA AGT    2323
Asn Ser Gly Gly Ala Ile Gln Ser Leu Glu Phe Asp Asp Asn Ala Ser
                730                 735                 740
TTG GTC AAG ATT GGG GTG AGA GGT TGC GGG GAG ATG AGC GTG TTT GCG    2371
Leu Val Lys Ile Gly Val Arg Gly Cys Gly Glu Met Ser Val Phe Ala
            745                 750                 755
TCT GAG AAA CCG GTT TGC TGC AAA ATT GAT GGG GTT AAG GTG AAA TTT    2419
Ser Glu Lys Pro Val Cys Cys Lys Ile Asp Gly Val Lys Val Lys Phe
        760                 765                 770
CTT TAT GAG GAC AAA ATG GCA AGA GTT CAA ATT CTG TGG CCT AGT TCT    2467
Leu Tyr Glu Asp Lys Met Ala Arg Val Gln Ile Leu Trp Pro Ser Ser
    775                 780                 785
TCA ACA TTG TCT TTG GTC CAG TTT TTA TTT TGA TCCCTAGGAA TCCTATGCAC  2520
Ser Thr Leu Ser Leu Val Gln Phe Leu Phe Stop
790                 795                 800
GTGTCTCTGT TTACAAGTAC TTTATATAAG TATAATATGT ATCTATTTCC ATTTTTAACT  2580
GTCTTTATGC AATTAGGTGG TCAATTAGTT ATTTGTTTGT GAAGTAACTA ACTTGCTTGT  2640
GTTGTAAGCT TATAATATAT GGTCAAGTTC CTCACTTGTA TATACCTGTT GTATGTATAA  2700
ATTTTACTAT ATATGACTAA CATCATTATC TTGTGAGCAA AAAAAA                 2746
```

(1) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 781 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: soybean (Glycine max)
(B) STRAIN: Williams 82
(F) TISSUE TYPE: seeds and leaves

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ala Pro Ser Ile Ser Lys Thr Val Glu Leu Asn Ser Phe Gly Leu
                5                   10                  15
Val Asn Gly Asn Leu Pro Leu Ser Ile Thr Leu Glu Gly Ser Asn Phe
            20                  25                  30
```

```
Leu Ala Asn Gly His Pro Phe Leu Thr Glu Val Pro Glu Asn Ile Ile
        35                  40                  45
Val Thr Pro Ser Pro Ile Asp Ala Lys Ser Ser Lys Asn Asn Glu Asp
    50                  55                  60
Asp Asp Val Val Gly Cys Phe Val Gly Phe His Ala Asp Glu Pro Arg
65                  70                  75                  80
Ser Arg His Val Ala Ser Leu Gly Lys Leu Arg Gly Ile Lys Phe Met
                85                  90                  95
Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp Val Gly Ser
            100                 105                 110
Asn Gly His Glu Leu Glu His Glu Thr Gln Met Met Leu Leu Asp Lys
        115                 120                 125
Asn Asp Gln Leu Gly Arg Pro Phe Val Leu Ile Leu Pro Ile Leu Gln
    130                 135                 140
Ala Ser Phe Arg Ala Ser Leu Gln Pro Gly Leu Asp Asp Tyr Val Asp
145                 150                 155                 160
Val Cys Met Glu Ser Gly Ser Thr Arg Val Cys Gly Ser Ser Phe Gly
                165                 170                 175
Ser Cys Leu Tyr Val His Val Gly His Asp Pro Tyr Gln Leu Leu Arg
            180                 185                 190
Glu Ala Thr Lys Val Val Arg Met His Leu Gly Thr Phe Lys Leu Leu
        195                 200                 205
Glu Glu Lys Thr Ala Pro Val Ile Ile Asp Lys Phe Gly Trp Cys Thr
    210                 215                 220
Trp Asp Ala Phe Tyr Leu Lys Val His Pro Ser Gly Val Trp Glu Gly
225                 230                 235                 240
Val Lys Gly Leu Val Glu Gly Gly Cys Pro Pro Gly Met Val Leu Ile
                245                 250                 255
Asp Asp Gly Trp Gln Ala Ile Cys His Asp Glu Asp Pro Ile Thr Asp
            260                 265                 270
Gln Glu Gly Met Lys Arg Thr Ser Ala Gly Glu Gln Met Pro Cys Arg
        275                 280                 285
Leu Val Lys Leu Glu Glu Asn Tyr Lys Phe Arg Gln Tyr Cys Ser Gly
    290                 295                 300
Lys Asp Ser Glu Lys Gly Met Gly Ala Phe Val Arg Asp Leu Lys Glu
305                 310                 315                 320
Gln Phe Arg Ser Val Glu Gln Val Tyr Val Trp His Ala Leu Cys Gly
                325                 330                 335
Tyr Trp Gly Gly Val Arg Pro Lys Val Pro Gly Met Pro Gln Ala Lys
            340                 345                 350
Val Val Thr Pro Lys Leu Ser Asn Gly Leu Lys Leu Thr Met Lys Asp
        355                 360                 365
Leu Ala Val Asp Lys Ile Val Ser Asn Gly Val Gly Leu Val Pro Pro
    370                 375                 380
His Leu Ala His Leu Leu Tyr Glu Gly Leu His Ser Arg Leu Glu Ser
385                 390                 395                 400
Ala Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu Met
                405                 410                 415
Leu Ser Glu Glu Tyr Gly Gly Arg Val Glu Leu Ala Lys Ala Tyr Tyr
            420                 425                 430
Lys Ala Leu Thr Ala Ser Val Lys Lys His Phe Lys Gly Asn Gly Val
        435                 440                 445
Ile Ala Ser Met Glu His Cys Asn Asp Phe Phe Leu Leu Gly Thr Glu
    450                 455                 460
Ala Ile Ala Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr Asp Pro
465                 470                 475                 480
Ser Gly Asp Pro Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met Val
                485                 490                 495
```

```
His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro Asp
            500                 505                 510
Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala Ala
        515                 520                 525
Ser Arg Ala Ile Ser Gly Gly Pro Val Tyr Val Ser Asp Cys Val Gly
    530                 535                 540
Lys His Asn Phe Lys Leu Leu Lys Ser Leu Ala Leu Pro Asp Gly Thr
545                 550                 555                 560
Ile Leu Arg Cys Gln His Tyr Ala Leu Pro Thr Arg Asp Cys Leu Phe
                565                 570                 575
Glu Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn Leu
            580                 585                 590
Asn Lys Tyr Thr Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly Gly
        595                 600                 605
Trp Cys Pro Val Thr Arg Arg Asn Lys Ser Ala Ser Glu Phe Ser Gln
    610                 615                 620
Thr Val Thr Cys Leu Ala Ser Pro Gln Asp Ile Glu Trp Ser Asn Gly
625                 630                 635                 640
Lys Ser Pro Ile Cys Ile Lys Gly Met Asn Val Phe Ala Val Tyr Leu
                645                 650                 655
Phe Lys Asp His Lys Leu Lys Leu Met Lys Ala Ser Glu Lys Leu Glu
            660                 665                 670
Val Ser Leu Glu Pro Phe Thr Phe Glu Leu Leu Thr Val Ser Pro Val
        675                 680                 685
Ile Val Leu Ser Lys Lys Leu Ile Gln Phe Ala Pro Ile Gly Leu Val
    690                 695                 700
Asn Met Leu Asn Thr Gly Gly Ala Ile Gln Ser Met Glu Phe Asp Asn
705                 710                 715                 720
His Ile Asp Val Val Lys Ile Gly Val Arg Gly Cys Gly Glu Met Lys
                725                 730                 735
Val Phe Ala Ser Glu Lys Pro Val Ser Cys Lys Leu Asp Gly Val Val
            740                 745                 750
Val Lys Phe Asp Tyr Glu Asp Lys Met Leu Arg Val Gln Val Pro Trp
        755                 760                 765
Pro Ser Ala Ser Lys Leu Ser Met Val Glu Phe Leu Phe Stop
    770                 775                 780
```

(1) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: peptide
(B) LOCATION: 62..2407
(C) IDENTIFICATION METHOD: by experiment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CCAAACCATA GCAAACCTAA GCACCAAACC TCTTTCTTTC AAGATCCTTG AATTCAGTCC    60
C ATG GCT CCA AGC ATA AGC AAA ACT GTG GAA CTA AAT TCA TTT GGT        106
  Met Ala Pro Ser Ile Ser Lys Thr Val Glu Leu Asn Ser Phe Gly
                  5                  10                  15
```

```
CTT GTC AAC GGT AAT TTG CCT TTG TCC ATA ACC CTA GAA GGA TCA AAT      154
Leu Val Asn Gly Asn Leu Pro Leu Ser Ile Thr Leu Glu Gly Ser Asn
                  20                  25                  30
TTC CTC GCC AAC GGC CAC CCT TTT CTC ACG GAA GTT CCC GAA AAC ATA      202
Phe Leu Ala Asn Gly His Pro Phe Leu Thr Glu Val Pro Glu Asn Ile
              35                  40                  45
ATA GTC ACC CCT TCA CCC ATC GAC GCC AAG AGT AGT AAG AAC AAC GAG      250
Ile Val Thr Pro Ser Pro Ile Asp Ala Lys Ser Ser Lys Asn Asn Glu
          50                  55                  60
GAC GAC GAC GTC GTA GGT TGC TTC GTG GGC TTC CAC GCG GAC GAG CCC      298
Asp Asp Asp Val Val Gly Cys Phe Val Gly Phe His Ala Asp Glu Pro
      65                  70                  75
AGA AGC CGA CAC GTG GCT TCC CTG GGG AAG CTC AGA GGA ATA AAA TTC      346
Arg Ser Arg His Val Ala Ser Leu Gly Lys Leu Arg Gly Ile Lys Phe
  80                  85                  90                  95
ATG AGC ATA TTC CGG TTT AAG GTG TGG TGG ACC ACT CAC TGG GTC GGT      394
Met Ser Ile Phe Arg Phe Lys Val Trp Trp Thr Thr His Trp Val Gly
                 100                 105                 110
AGC AAC GGA CAC GAA CTG GAG CAC GAG ACA CAG ATG ATG CTT CTC GAC      442
Ser Asn Gly His Glu Leu Glu His Glu Thr Gln Met Met Leu Leu Asp
             115                 120                 125
AAA AAC GAC CAG CTC GGA CGC CCC TTT GTG TTG ATT CTC CCG ATC CTC      490
Lys Asn Asp Gln Leu Gly Arg Pro Phe Val Leu Ile Leu Pro Ile Leu
         130                 135                 140
CAA GCC TCG TTC CGA GCC TCC CTG CAA CCC GGT TTG GAT GAT TAC GTG      538
Gln Ala Ser Phe Arg Ala Ser Leu Gln Pro Gly Leu Asp Asp Tyr Val
     145                 150                 155
GAC GTT TGC ATG GAG AGC GGG TCG ACA CGT GTC TGT GGC TCC AGC TTC      586
Asp Val Cys Met Glu Ser Gly Ser Thr Arg Val Cys Gly Ser Ser Phe
160                 165                 170                 175
GGG AGC TGC TTA TAC GTC CAC GTT GGC CAT GAC CCG TAT CAG TTG CTT      634
Gly Ser Cys Leu Tyr Val His Val Gly His Asp Pro Tyr Gln Leu Leu
                 180                 185                 190
AGA GAA GCA ACT AAA GTC GTT AGG ATG CAT TTG GGG ACG TTC AAG CTT      682
Arg Glu Ala Thr Lys Val Val Arg Met His Leu Gly Thr Phe Lys Leu
             195                 200                 205
CTC GAG GAG AAA ACC GCG CCA GTG ATC ATA GAC AAG TTT GGT TGG TGT      730
Leu Glu Glu Lys Thr Ala Pro Val Ile Ile Asp Lys Phe Gly Trp Cys
         210                 215                 220
ACA TGG GAC GCG TTT TAC TTG AAG GTG CAT CCC TCA GGT GTG TGG GAA      778
Thr Trp Asp Ala Phe Tyr Leu Lys Val His Pro Ser Gly Val Trp Glu
     225                 230                 235
GGG GTG AAA GGG TTG GTG GAG GGA GGG TGC CCT CCA GGG ATG GTC CTA      826
Gly Val Lys Gly Leu Val Glu Gly Gly Cys Pro Pro Gly Met Val Leu
240                 245                 250                 255
ATC GAC GAC GGG TGG CAA GCC ATT TGT CAC GAC GAG GAC CCC ATA ACG      874
Ile Asp Asp Gly Trp Gln Ala Ile Cys His Asp Glu Asp Pro Ile Thr
                 260                 265                 270
GAC CAA GAG GGT ATG AAG CGA ACC TCC GCA GGG GAG CAA ATG CCA TGC      922
Asp Gln Glu Gly Met Lys Arg Thr Ser Ala Gly Glu Gln Met Pro Cys
             275                 280                 285
AGG TTG GTG AAG TTG GAG GAA AAT TAC AAG TTC AGA CAG TAT TGT AGT      970
Arg Leu Val Lys Leu Glu Glu Asn Tyr Lys Phe Arg Gln Tyr Cys Ser
         290                 295                 300
GGA AAG GAT TCT GAG AAG GGT ATG GGT GCC TTT GTT AGG GAC TTG AAG     1018
Gly Lys Asp Ser Glu Lys Gly Met Gly Ala Phe Val Arg Asp Leu Lys
     305                 310                 315
```

```
GAA CAG TTT AGG AGC GTG GAG CAG GTG TAT GTG TGG CAC GCG CTT TGT    1066
Glu Gln Phe Arg Ser Val Glu Gln Val Tyr Val Trp His Ala Leu Cys
320                 325                 330                 335
GGG TAT TGG GGT GGG GTC AGA CCC AAG GTT CCG GGC ATG CCC CAG GCT    1114
Gly Tyr Trp Gly Gly Val Arg Pro Lys Val Pro Gly Met Pro Gln Ala
                340                 345                 350
AAG GTT GTC ACT CCG AAG CTG TCC AAT GGA CTA AAA TTG ACA ATG AAG    1162
Lys Val Val Thr Pro Lys Leu Ser Asn Gly Leu Lys Leu Thr Met Lys
            355                 360                 365
GAT TTA GCG GTG GAT AAG ATC GTC AGT AAC GGA GTT GGA CTG GTG CCA    1210
Asp Leu Ala Val Asp Lys Ile Val Ser Asn Gly Val Gly Leu Val Pro
        370                 375                 380
CCA CAC CTG GCT CAC CTT TTG TAC GAG GGG CTC CAC TCC CGT TTG GAA    1258
Pro His Leu Ala His Leu Leu Tyr Glu Gly Leu His Ser Arg Leu Glu
    385                 390                 395
TCT GCG GGT ATT GAC GGT GTT AAG GTT GAC GTT ATA CAC TTG CTC GAG    1306
Ser Ala Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu Glu
400                 405                 410                 415
ATG CTA TCC GAG GAA TAC GGT GGC CGT GTT GAG CTA GCC AAA GCT TAT    1354
Met Leu Ser Glu Glu Tyr Gly Gly Arg Val Glu Leu Ala Lys Ala Tyr
                420                 425                 430
TAC AAA GCG CTC ACT GCT TCG GTG AAG AAG CAT TTC AAA GGC AAT GGG    1402
Tyr Lys Ala Leu Thr Ala Ser Val Lys Lys His Phe Lys Gly Asn Gly
            435                 440                 445
GTC ATT GCG AGC ATG GAG CAT TGT AAT GAC TTC TTT CTC CTT GGT ACC    1450
Val Ile Ala Ser Met Glu His Cys Asn Asp Phe Phe Leu Leu Gly Thr
        450                 455                 460
GAA GCC ATA GCC CTT GGG CGC GTA GGA GAT GAT TTT TGG TGC ACT GAT    1498
Glu Ala Ile Ala Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr Asp
    465                 470                 475
CCC TCT GGA GAT CCA AAT GGC ACG TAT TGG CTC CAA GGG TGT CAC ATG    1546
Pro Ser Gly Asp Pro Asn Gly Thr Tyr Trp Leu Gln Gly Cys His Met
480                 485                 490                 495
GTG CAC TGT GCC TAC AAC AGC TTG TGG ATG GGG AAT TTT ATT CAG CCG    1594
Val His Cys Ala Tyr Asn Ser Leu Trp Met Gly Asn Phe Ile Gln Pro
                500                 505                 510
GAT TGG GAC ATG TTC CAG TCC ACT CAC CCT TGT GCC GAA TTC CAT GC     1642
Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His Ala
            515                 520                 525
GCC TCT AGG GCC ATC TCT GGT GGA CCA GTT TAC GTT AGT GAT TGT GTT    1690
Ala Ser Arg Ala Ile Ser Gly Gly Pro Val Tyr Val Ser Asp Cys Val
        530                 535                 540
GGA AAG CAC AAC TTC AAG TTG CTC AAG AGC CTC GCT TTG CCT GAT GGG    1738
Gly Lys His Asn Phe Lys Leu Leu Lys Ser Leu Ala Leu Pro Asp Gly
    545                 550                 555
ACG ATT TTG CGT TGT CAA CAC TAT GCA CTC CCC ACA CGA GAC TGT TTG    1786
Thr Ile Leu Arg Cys Gln His Tyr Ala Leu Pro Thr Arg Asp Cys Leu
560                 565                 570                 575
TTT GAA GAC CCC TTG CAT GAT GGG AAG ACA ATG CTC AAA ATT TGG AAT    1834
Phe Glu Asp Pro Leu His Asp Gly Lys Thr Met Leu Lys Ile Trp Asn
                580                 585                 590
CTC AAC AAA TAT ACA GGT GTT TTG GGT CTA TTT AAT TGC CAA GGA GGT    1882
Leu Asn Lys Tyr Thr Gly Val Leu Gly Leu Phe Asn Cys Gln Gly Gly
            595                 600                 605
GGG TGG TGT CCC GTA ACT AGG AGA AAC AAG AGT GCC TCT GAA TTT TCA    1930
Gly Trp Cys Pro Val Thr Arg Arg Asn Lys Ser Ala Ser Glu Phe Ser
        610                 615                 620
```

```
CAA ACT GTG ACA TGC TTA GCG AGT CCT CAA GAC ATT GAA TGG AGC AAT    1978
Gln Thr Val Thr Cys Leu Ala Ser Pro Gln Asp Ile Glu Trp Ser Asn
    625                 630                 635
GGG AAA AGC CCA ATA TGC ATA AAA GGG ATG AAT GTG TTT GCT GTA TAT    2026
Gly Lys Ser Pro Ile Cys Ile Lys Gly Met Asn Val Phe Ala Val Tyr
640                 645                 650                 655
TTG TTC AAG GAC CAC AAA CTA AAG CTC ATG AAG GCA TCA GAG AAA TTG    2074
Leu Phe Lys Asp His Lys Leu Lys Leu Met Lys Ala Ser Glu Lys Leu
                660                 665                 670
GAA GTT TCA CTT GAG CCA TTT ACT TTT GAG CTA TTG ACA GTG TCT CCA    2122
Glu Val Ser Leu Glu Pro Phe Thr Phe Glu Leu Leu Thr Val Ser Pro
            675                 680                 685
GTG ATT GTG CTG TCA AAA AAG TTA ATT CAA TTT GCT CCA ATT GGA TTA    2170
Val Ile Val Leu Ser Lys Lys Leu Ile Gln Phe Ala Pro Ile Gly Leu
        690                 695                 700
GTG AAC ATG CTT AAC ACT GGT GGT GCC ATT CAG TCC ATG GAG TTT GAC    2218
Val Asn Met Leu Asn Thr Gly Gly Ala Ile Gln Ser Met Glu Phe Asp
    705                 710                 715
AAC CAC ATA GAT GTG GTC AAA ATT GGG GTT AGG GGT TGT GGG GAG ATG    2266
Asn His Ile Asp Val Val Lys Ile Gly Val Arg Gly Cys Gly Glu Met
720                 725                 730                 735
AAG GTG TTT GCA TCA GAG AAA CCA GTT AGT TGC AAA CTA GAT GGG GTA    2314
Lys Val Phe Ala Ser Glu Lys Pro Val Ser Cys Lys Leu Asp Gly Val
                740                 745                 750
GTT GTA AAA TTT GAT TAT GAG GAT AAA ATG CTG AGA GTG CAA GTT CCC    2362
Val Val Lys Phe Asp Tyr Glu Asp Lys Met Leu Arg Val Gln Val Pro
            755                 760                 765
TGG CCT AGT GCT TCA AAA TTG TCA ATG GTT GAG TTT TTA TTT TGA TCCCT  2412
Trp Pro Ser Ala Ser Lys Leu Ser Met Val Glu Phe Leu Phe Stop
        770                 775                 780
GAAGGTGAAT TTGGGATACT ATGATGTTTG ACTCTCTTTT TAAGTAATAA GAGTCATATT  2472
TTTCTGTTGT AAAAAAAAAA AAAAAA                                       2498
```

(1) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 587 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Japanese artichoke (Stachys sieboldii)
(F) TISSUE TYPE: leaves

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Thr Asn Gly Ser Asp Leu Glu Arg Glu Thr Gln Ile Val Val Leu Asp
  1               5                  10                  15
Lys Ser Asp Asp Arg Pro Tyr Ile Val Leu Leu Pro Leu Ile Glu Gly
            20                  25                  30
Gln Phe Arg Ala Ser Leu Gln Pro Gly Val Asp Asp Phe Ile Asp Ile
        35                  40                  45
Cys Val Glu Ser Gly Ser Thr Lys Val Asn Glu Ser Ser Phe Arg Ala
    50                  55                  60
Ser Leu Tyr Met His Ala Gly Asp Asp Pro Phe Thr Leu Val Lys Asp
65                  70                  75                  80
```

```
Ala Val Lys Val Ala Arg His His Leu Gly Thr Phe Arg Leu Leu Glu
                  85                  90                  95
Glu Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp Cys Thr Trp
             100                 105                 110
Asp Ala Phe Tyr Leu Asn Val Gln Pro His Gly Val Met Glu Gly Val
         115                 120                 125
Gln Gly Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val Leu Ile Asp
     130                 135                 140
Asp Gly Trp Gln Ser Ile Cys His Asp Asn Asp Ala Leu Thr Thr Glu
145                 150                 155                 160
Gly Met Gly Arg Thr Ser Ala Gly Glu Gln Met Pro Cys Arg Leu Ile
                 165                 170                 175
Lys Phe Glu Glu Asn Tyr Lys Phe Arg Glu Tyr Glu Ser Pro Asn Lys
             180                 185                 190
Thr Gly Pro Gly Pro Asn Thr Gly Met Gly Ala Phe Ile Arg Asp Met
         195                 200                 205
Lys Asp Asn Phe Lys Ser Val Asp Tyr Val Tyr Val Trp His Ala Leu
     210                 215                 220
Cys Gly Tyr Trp Gly Gly Leu Arg Pro Asn Val Pro Gly Leu Pro Glu
225                 230                 235                 240
Ala Lys Leu Ile Glu Pro Lys Leu Thr Pro Gly Leu Lys Thr Thr Met
                 245                 250                 255
Glu Asp Leu Ala Val Asp Lys Ile Val Asn Asn Gly Val Gly Leu Val
             260                 265                 270
Pro Pro Glu Phe Val Glu Gln Met Tyr Glu Gly Leu His Ser His Leu
         275                 280                 285
Glu Ser Val Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu Leu
     290                 295                 300
Glu Met Leu Cys Glu Asp Tyr Gly Gly Arg Val Asp Leu Ala Lys Ala
305                 310                 315                 320
Tyr Tyr Lys Ala Leu Ser Ser Ser Val Asn Asn His Phe Asn Gly Asn
                 325                 330                 335
Gly Val Ile Ala Gly Leu Glu His Cys Asn Asp Phe Met Phe Leu Gly
             340                 345                 350
Thr Glu Ala Ile Thr Leu Gly Arg Val Gly Asp Asp Phe Trp Cys Thr
         355                 360                 365
Asp Pro Ser Gly Asp Pro Asn Gly Thr Phe Trp Leu Gln Gly Cys His
     370                 375                 380
Met Val His Cys Ala Tyr Asn Ser Ile Trp Met Gly Asn Phe Ile His
385                 390                 395                 400
Pro Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe His
                 405                 410                 415
Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Ser
             420                 425                 430
Val Gly Lys His Asn Phe Glu Leu Leu Arg Ser Leu Val Leu Pro Asp
         435                 440                 445
Gly Ser Ile Leu Arg Cys Asp Tyr Tyr Ala Leu Pro Thr Arg Asp Cys
     450                 455                 460
Leu Phe Glu Asp Pro Leu His Asn Gly Lys Thr Met Leu Lys Ile Trp
465                 470                 475                 480
Asn Tyr Asn Lys Phe Thr Gly Val Val Gly Thr Phe Asn Cys Gln Gly
                 485                 490                 495
Gly Gly Trp Ser Arg Glu Val Arg Arg Asn Gln Cys Ala Ala Glu Tyr
             500                 505                 510
Ser His Ala Val Ser Ser Ser Ala Gly Pro Ser Asp Ile Glu Trp Lys
         515                 520                 525
Gln Gly Thr Ser Pro Ile Asp Val Asp Gly Val Lys Thr Phe Ala Leu
     530                 535                 540
```

```
Tyr Leu Phe His Glu Lys Lys Leu Val Leu Ser Lys Pro Ser Asp Lys
545                 550                 555                 560
Ile Asp Ile Thr Leu Glu Pro Phe Asp Phe Glu Leu Ile Thr Val Ser
                565                 570                 575
Pro Val Lys Thr Leu Ala Asn Cys Thr Val Gln
            580                 585
```

(1) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: peptide
(B) LOCATION: 2..1762
(C) IDENTIFICATION METHOD: by experiment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
G ACA AAC GGG TCG GAT CTT GAG CGG GAA ACT CAA ATA GTC GTG CTC        46
  Thr Asn Gly Ser Asp Leu Glu Arg Glu Thr Gln Ile Val Val Leu
    1               5                  10                  15
GAC AAG TCC GAC GAC AGG CCC TAC ATC GTG CTG CTT CCG CTC ATC GAG      94
Asp Lys Ser Asp Asp Arg Pro Tyr Ile Val Leu Leu Pro Leu Ile Glu
                20                  25                  30
GGG CAG TTT CGG GCT TCC CTT CAG CCC GGT GTG GAT GAT TTT ATC GAT     142
Gly Gln Phe Arg Ala Ser Leu Gln Pro Gly Val Asp Asp Phe Ile Asp
            35                  40                  45
ATT TGT GTC GAA AGC GGG TCA ACC AAG GTC AAC GAG TCC TCG TTC CGT     190
Ile Cys Val Glu Ser Gly Ser Thr Lys Val Asn Glu Ser Ser Phe Arg
        50                  55                  60
GCT TCG CTC TAC ATG CAC GCC GGT GAT GAC CCT TTT ACC CTG GTG AAG     238
Ala Ser Leu Tyr Met His Ala Gly Asp Asp Pro Phe Thr Leu Val Lys
    65                  70                  75
GAC GCC GTG AAG GTG GCG CGC CAC CAC CTC GGG ACG TTC AGG CTG CTG     286
Asp Ala Val Lys Val Ala Arg His His Leu Gly Thr Phe Arg Leu Leu
80                  85                  90                  95
GAG GAG AAA ACT CCG CCG GGG ATC GTC GAC AAA TTC GGG TGG TGC ACG     334
Glu Glu Lys Thr Pro Pro Gly Ile Val Asp Lys Phe Gly Trp Cys Thr
                100                 105                 110
TGG GAT GCG TTC TAC CTC AAC GTC CAG CCC CAC GGC GTT ATG GAG GGC     382
Trp Asp Ala Phe Tyr Leu Asn Val Gln Pro His Gly Val Met Glu Gly
            115                 120                 125
GTG CAG GGG CTG GTT GAC GGC GGA TGT CCG CCG GGG CTG GTG TTG ATC     430
Val Gln Gly Leu Val Asp Gly Gly Cys Pro Pro Gly Leu Val Leu Ile
        130                 135                 140
GAC GAC GGG TGG CAG TCC ATT TGT CAC GAC AAC GAC GCG CTC ACC ACC     478
Asp Asp Gly Trp Gln Ser Ile Cys His Asp Asn Asp Ala Leu Thr Thr
    145                 150                 155
GAG GGG ATG GGG AGA ACC TCC GCC GGA GAG CAA ATG CCC TGC AGG TTG     526
Glu Gly Met Gly Arg Thr Ser Ala Gly Glu Gln Met Pro Cys Arg Leu
160                 165                 170                 175
```

```
ATC AAG TTT GAG GAG AAT TAC AAG TTC AGG GAG TAC GAG AGC CCG AAT      574
Ile Lys Phe Glu Glu Asn Tyr Lys Phe Arg Glu Tyr Glu Ser Pro Asn
                180                 185                 190
AAA ACT GGG CCG GGC CCG AAT ACG GGG ATG GGG GCC TTT ATT CGT GAC      622
Lys Thr Gly Pro Gly Pro Asn Thr Gly Met Gly Ala Phe Ile Arg Asp
            195                 200                 205
ATG AAG GAC AAT TTC AAG AGT GTG GAC TAC GTG TAC GTG TGG CAT GCG      670
Met Lys Asp Asn Phe Lys Ser Val Asp Tyr Val Tyr Val Trp His Ala
        210                 215                 220
TTG TGT GGT TAT TGG GGC GGG CTC AGG CCC AAT GTT CCG GGC CTG CCC      718
Leu Cys Gly Tyr Trp Gly Gly Leu Arg Pro Asn Val Pro Gly Leu Pro
    225                 230                 235
GAG GCT AAG CTC ATT GAG CCC AAA CTG ACT CCT GGG CTT AAG ACC ACC      766
Glu Ala Lys Leu Ile Glu Pro Lys Leu Thr Pro Gly Leu Lys Thr Thr
240                 245                 250                 255
ATG GAA GAT TTG GCT GTT GAT AAG ATT GTC AAC AAT GGC GTG GGT CTG      814
Met Glu Asp Leu Ala Val Asp Lys Ile Val Asn Asn Gly Val Gly Leu
                260                 265                 270
GTC CCA CCG GAG TTT GTT GAA CAA ATG TAT GAA GGA TTA CAT TCA CAT      862
Val Pro Pro Glu Phe Val Glu Gln Met Tyr Glu Gly Leu His Ser His
            275                 280                 285
CTC GAA TCT GTG GGG ATT GAT GGA GTC AAA GTT GAC GTC ATC CAT TTG      910
Leu Glu Ser Val Gly Ile Asp Gly Val Lys Val Asp Val Ile His Leu
        290                 295                 300
TTG GAA ATG TTG TGT GAA GAC TAT GGT GGG AGA GTG GAC TTA GCC AAG      958
Leu Glu Met Leu Cys Glu Asp Tyr Gly Gly Arg Val Asp Leu Ala Lys
    305                 310                 315
GCT TAT TAC AAG GCC TTA TCA AGC TCA GTT AAC AAC CAC TTC AAC GGC     1006
Ala Tyr Tyr Lys Ala Leu Ser Ser Ser Val Asn Asn His Phe Asn Gly
320                 325                 330                 335
AAC GGC GTC ATC GCT GGC CTG GAG CAC TGC AAT GAC TTC ATG TTT CTC     1054
Asn Gly Val Ile Ala Gly Leu Glu His Cys Asn Asp Phe Met Phe Leu
                340                 345                 350
GGA ACC GAG GCC ATT ACC TTG GGT CGT GTC GGG GAT GAT TTT TGG TGC     1102
Gly Thr Glu Ala Ile Thr Leu Gly Arg Val Gly Asp Asp Phe Trp Cys
            355                 360                 365
ACT GAT CCA TCT GGA GAT CCC AAT GGC ACG TTC TGG TTG CAA GGG TGT     1150
Thr Asp Pro Ser Gly Asp Pro Asn Gly Thr Phe Trp Leu Gln Gly Cys
        370                 375                 380
CAC ATG GTG CAC TGC GCC TAC AAC AGC ATA TGG ATG GGT AAT TTC ATC     1198
His Met Val His Cys Ala Tyr Asn Ser Ile Trp Met Gly Asn Phe Ile
    385                 390                 395
CAC CCT GAT TGG GAC ATG TTT CAA TCG ACT CAC CCT TGC GCT GAA TTC     1246
His Pro Asp Trp Asp Met Phe Gln Ser Thr His Pro Cys Ala Glu Phe
400                 405                 410                 415
CAC GCT GCC TCA CGA GCC ATC TCC GGC GGG CCC ATT TAC GTC AGT GAC     1294
His Ala Ala Ser Arg Ala Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp
                420                 425                 430
TCG GTC GGA AAG CAC AAC TTC GAG CTC CTT AGG AGC CTC GTT CTT CCC     1342
Ser Val Gly Lys His Asn Phe Glu Leu Leu Arg Ser Leu Val Leu Pro
            435                 440                 445
GAT GGC TCC ATC CTC CGT TGT GAT TAC TAC GCG CTT CCG ACT CGC GAT     1390
Asp Gly Ser Ile Leu Arg Cys Asp Tyr Tyr Ala Leu Pro Thr Arg Asp
        450                 455                 460
TGC CTC TTT GAA GAT CCA CTT CAC AAT GGC AAG ACT ATG CTC AAA ATT     1438
Cys Leu Phe Glu Asp Pro Leu His Asn Gly Lys Thr Met Leu Lys Ile
    465                 470                 475
```

```
TGG AAT TAT AAC AAG TTC ACC GGA GTT GTC GGA ACT TTC AAC TGC CAA    1486
Trp Asn Tyr Asn Lys Phe Thr Gly Val Val Gly Thr Phe Asn Cys Gln
480                 485                 490                 495
GGT GGC GGG TGG AGC CGG GAA GTG CGT CGC AAC CAA TGC GCT GCC GAG    1534
Gly Gly Gly Trp Ser Arg Glu Val Arg Arg Asn Gln Cys Ala Ala Glu
                500                 505                 510
TAT TCC CAC GCC GTC TCC TCT AGC GCT GGT CCG AGT GAC ATT GAG TGG    1582
Tyr Ser His Ala Val Ser Ser Ser Ala Gly Pro Ser Asp Ile Glu Trp
            515                 520                 525
AAG CAA GGA ACG AGT CCG ATC GAC GTC GAC GGC GTC AAA ACA TTC GCG    1630
Lys Gln Gly Thr Ser Pro Ile Asp Val Asp Gly Val Lys Thr Phe Ala
        530                 535                 540
TTG TAC CTA TTC CAC GAG AAG AAA CTC GTC CTT TCT AAG CCA TCA GAC    1678
Leu Tyr Leu Phe His Glu Lys Lys Leu Val Leu Ser Lys Pro Ser Asp
    545                 550                 555
AAA ATC GAC ATC ACG CTT GAG CCC TTC GAT TTT GAG CTG ATA ACC GTT    1726
Lys Ile Asp Ile Thr Leu Glu Pro Phe Asp Phe Glu Leu Ile Thr Val
560                 565                 570                 575
TCT CCA GTC AAA ACT CTA GCC AAT TGC ACC GTC CAA                    1762
Ser Pro Val Lys Thr Leu Ala Asn Cys Thr Val Gln
                580                 585
```

(1) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(vi) ORIGINAL SOURCE:
(A) ORGANISM: corn (Zea mays L.)
(B) STRAIN: Pioneer 3358
(F) TISSUE TYPE: leaves

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Gln Ser Thr His Pro Cys Ala Ala Phe His Ala Ala Ser Arg Ala Ile
                5                  10                  15
Ser Gly Gly Pro Ile Tyr Val Ser Asp Ser Val Gly Gln His Asp Phe
            20                  25                  30
Ala Leu Leu Arg Arg Leu Ala Leu Pro Asp Gly Thr Val Leu Arg Cys
        35                  40                  45
Glu Gly His Ala Leu Pro Thr Arg Asp Cys Leu Phe Ala Asp Pro Leu
    50                  55                  60
His Asp Gly Arg Thr Val Leu Lys Ile Trp Asn Val Asn Arg Phe Ala
65                  70                  75                  80
Gly Val Val Gly Ala Phe Asn Cys Gln Gly Gly Gly Trp Ser Pro Glu
                85                  90                  95
Ala Arg Arg Asn Lys Cys Phe Ser Glu Phe Ser Val Pro Leu Ala Ala
            100                 105                 110
Arg Ala Ser Pro Ser Asp Val Glu Trp Lys Ser Gly Lys Ala Gly Pro
        115                 120                 125
Gly Val Ser Val Lys Asp Val Ser Gln Phe Ala Val Tyr Ala Val Glu
    130                 135                 140
Ala Arg Thr Leu Gln Leu Leu Arg Pro Asp Glu Gly Val Asp Leu Thr
145                 150                 155                 160
```

```
Leu Gln Pro Phe Thr Tyr Glu Leu Phe Val Val Ala Pro Val Arg Val
                165                 170                 175
Ile Ser His Glu Arg Ala Ile Lys Phe Ala Pro Ile Gly Leu Ala Asn
            180                 185                 190
Met Leu Asn Thr Ala Gly Ala Val Gln Ala Phe Glu Ala Lys Lys Asp
        195                 200                 205
Ala Ser Gly Val Thr Ala Glu Val Phe Val Lys Gly Ala Gly Glu Leu
    210                 215                 220
Val Ala Tyr Ser Ser Ala Thr Pro Arg Leu Cys Lys Val Asn Gly Asp
225                 230                 235                 240
Glu Ala Glu Phe Thr Tyr Lys Asp Gly Val Val Thr Val Asp Val Pro
                245                 250                 255
Trp Ser Gly Ser Ser Ser Lys Leu Cys Cys Val Gln Tyr Val Tyr Stop
            260                 265                 270
```

(1) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: peptide
(B) LOCATION: 2..817
(C) IDENTIFICATION METHOD: by experiment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
C CAG TCC ACG CAC CCC TGC GCC GCC TTC CAC GCC GCG TCC CGC GCC        46
  Gln Ser Thr His Pro Cys Ala Ala Phe His Ala Ala Ser Arg Ala
                  5                  10                  15
ATC TCC GGC GGG CCC ATC TAC GTC AGC GAC TCG GTG GGG CAG CAC GAC      94
Ile Ser Gly Gly Pro Ile Tyr Val Ser Asp Ser Val Gly Gln His Asp
              20                  25                  30
TTC GCG CTG CTC CGC CGC CTG GCG CTC CCC GAC GGC ACC GTC CTC CGG     142
Phe Ala Leu Leu Arg Arg Leu Ala Leu Pro Asp Gly Thr Val Leu A
          35                  40                  45
TGC GAG GGC CAC GCG CTG CCC ACG CGC GAC TGC CTC TTC GCC GAC CCG     190
Cys Glu Gly His Ala Leu Pro Thr Arg Asp Cys Leu Phe Ala Asp Pro
      50                  55                  60
CTC CAC GAC GGC CGG ACC GTG CTC AAG ATC TGG AAC GTG AAC CGC TTC     238
Leu His Asp Gly Arg Thr Val Leu Lys Ile Trp Asn Val Asn Arg Phe
  65                  70                  75
GCC GGC GTC GTC GGC GCC TTC AAC TGC CAG GGC GGC GGG TGG AGC CCC     286
Ala Gly Val Val Gly Ala Phe Asn Cys Gln Gly Gly Gly Trp Ser Pro
80                  85                  90                  95
GAG GCG CGG CGG AAC AAG TGC TTC TCG GAG TTC TCC GTG CCC CTG GCC     334
Glu Ala Arg Arg Asn Lys Cys Phe Ser Glu Phe Ser Val Pro Leu Ala
              100                 105                 110
GCG CGC GCC TCG CCG TCC GAC GTC GAG TGG AAG AGC GGC AAG GCG GGG     382
Ala Arg Ala Ser Pro Ser Asp Val Glu Trp Lys Ser Gly Lys Ala Gly
          115                 120                 125
```

```
CCA GGC GTC AGC GTC AAG GAC GTC TCC CAG TTC GCC GTG TAC GCG GTC      430
Pro Gly Val Ser Val Lys Asp Val Ser Gln Phe Ala Val Tyr Ala Val
        130                 135                 140
GAG GCC AGG ACG CTG CAG CTG CTG CGC CCC GAC GAG GGC GTC GAC CTC      478
Glu Ala Arg Thr Leu Gln Leu Leu Arg Pro Asp Glu Gly Val Asp Leu
    145                 150                 155
ACG CTG CAG CCC TTC ACC TAC GAG CTC TTC GTC GTT GCC CCC GTG CGC      526
Thr Leu Gln Pro Phe Thr Tyr Glu Leu Phe Val Val Ala Pro Val Arg
160                 165                 170                 175
GTC ATC TCG CAT GAG CGG GCC ATC AAG TTC GCG CCC ATC GGA CTC GCC      574
Val Ile Ser His Glu Arg Ala Ile Lys Phe Ala Pro Ile Gly Leu Ala
                180                 185                 190
AAC ATG CTC AAC ACC GCC GGC GCC GTG CAG GCG TTC GAG GCC AAG AAA      622
Asn Met Leu Asn Thr Ala Gly Ala Val Gln Ala Phe Glu Ala Lys Lys
            195                 200                 205
GAT GCT AGC GGC GTC ACG GCA GAG GTG TTC GTG AAG GGC GCA GGG GAG      670
Asp Ala Ser Gly Val Thr Ala Glu Val Phe Val Lys Gly Ala Gly Glu
        210                 215                 220
CTG GTG GCG TAC TCG TCG GCG ACG CCC AGG CTC TGC AAG GTG AAC GGC      718
Leu Val Ala Tyr Ser Ser Ala Thr Pro Arg Leu Cys Lys Val Asn Gly
    225                 230                 235
GAC GAG GCC GAG TTC ACG TAC AAG GAC GGC GTG GTC ACC GTC GAC GTG      766
Asp Glu Ala Glu Phe Thr Tyr Lys Asp Gly Val Val Thr Val Asp Val
240                 245                 250                 255
CCG TGG TCG GGG TCG TCG TCG AAG CTG TGT TGC GTC CAG TAC GTC TAC      814
Pro Trp Ser Gly Ser Ser Ser Lys Leu Cys Cys Val Gln Tyr Val Tyr
                260                 265                 270
TGA GCCGGACGGG CCGATGACTC TGCGTCTCTG CTCCCTGCTG GCCTGCTCAG GAC      873
Stop
ATAATCTAAT GTTTAGAGCT TACCAGGTTT TACAGCTCTA TCAGTTTACT TTTGTTTTTC   933
TGCTCTTCGT TTTTTAAGAA TTATTTCTAT TGTGTGAATT AATGAGTGCT TTCCTTCTAA   993
AAA                                                                 996
```

Claims

1. A nucleic acid molecule encoding a plant raffinose synthetase, said synthetase being capable of producing raffinose by combining a D-galactosyl group through an $\alpha(1\rightarrow6)$ bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule, said nucleic acid molecule being selected from the group consisting of:

   (a) nucleic acid molecules encoding a protein with the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO:7;
   (b) nucleic acid molecules comprising the coding region of the nucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8;
   (c) nucleic acid molecules hybridizing to the complementary strand of a nucleic acid molecule indicated under (a) or (b);
   (d) nucleic acid molecules the sequence of which differs from the sequence of a nucleic acid molecule of any one of (a) to (c) due to the degeneracy of the genetic code; and
   (e) a fragment of any of the nucleic acid molecules of any one of (a) to (d).

2. The nucleic acid molecule encoding a plant raffinose synthetase according to claim 1, wherein the plant is a dicotelydon, preferably a leguminous plant, preferably a broad bean or soybean, or a lamiaceous plant, preferably a Japanese artichoke, or a monocotyledon, preferably a gramineous plant, preferably corn.

3. A probe or primer that specifically hybridizes to a nucleic acid molecul of claim 1 or 2.

4. The probe or primer of claim 3 that has a length of at least 15, 17, 21 or 50 nucleotides.

5. A chimeric gene comprising th nucleic acid molecule of claim 1 or 2, or the probe or primer of claim 3 or 4.

6. The chimeric gene of claim 5 functionally linked to a promoter.

7. A plasmid comprising a nucleic acid molecule of claim 1 or 2, optionally in functional combination with a promoter, or a chimeric gene of claim 5 or 6.

8. A host organism containing the chimeric gene of claim 5 or 6, or the plasmid of claim 7.

9. The host organism of claim 8 which is a microorganism, a plant cell or a plant.

10. A method for the production of a raffinose synthetase protein or a portion thereof, comprising the steps of isolating and purifying a raffinose synthetase protein or a portion thereof from a culture obtained by cultivating the host organism of claim 8 or 9.

11. A raffinose synthetase protein encoded by the nucleic acid molecule of claim 1 or 2 or obtained by the method of claim 10.

12. An antisense nucleic acid molecule or a ribozyme specifically hybridizing or binding to a nucleic acid molecule of claim 1 or 2.

13. A method for modifying metabolism of a derived host organism, which comprises introducing the nucleic acid molecule, probe, primer or chimeric gene of any one of claims 1 to 6 or an antisense nucleic acid molecule or ribozyme of claim 12 into said host organism or a cell thereof, so that the content of the raffinose family oligosaccharides in said host organism or cell thereof is changed.

14. A raffinose synthetase protein related to the raffinose synthetase protein of claim 11 by deletion, replacement, modification or addition of one or several amino acids.

15. An anti-raffinose synthetase antibody capable of binding to the raffinose synthetase protein of claim 11 or 14.

16. Use of the anti-raffinose synthetase antibody of claim 15 for the detection of a raffinose synthetase protein.

17. A method for the detection of a raffinose synthetase gene or a fragment thereof which comprises hybridizing a probe of claim 3 or 4 to an organism-derived genomic DNA or cDNA fragment, and detecting the presence of hybridization.

18. A method for the amplification of a raffinose synthetase gene or a part thereof, which comprises annealing the primer of claim 3 or 4 to organism-derived genomic DNA or cDNA, and amplifying the resulting DNA fragment.

19. The method of claim 18 wherein the DNA fragment is amplified by polymerase chain reaction (PCR).

20. The method of any one of claims 17 to 19, wherein the organism is a plant.

21. A method for obtaining a raffinose synthetase gene, comprising the steps of identifying a DNA fragment containing a raffinose synthetase gene or a gene fragment thereof by the method of any one of claims 17 to 20, and isolating and purifying the DNA fragment identified.

22. A raffinose synthetase gene obtained by the method of claim 21.

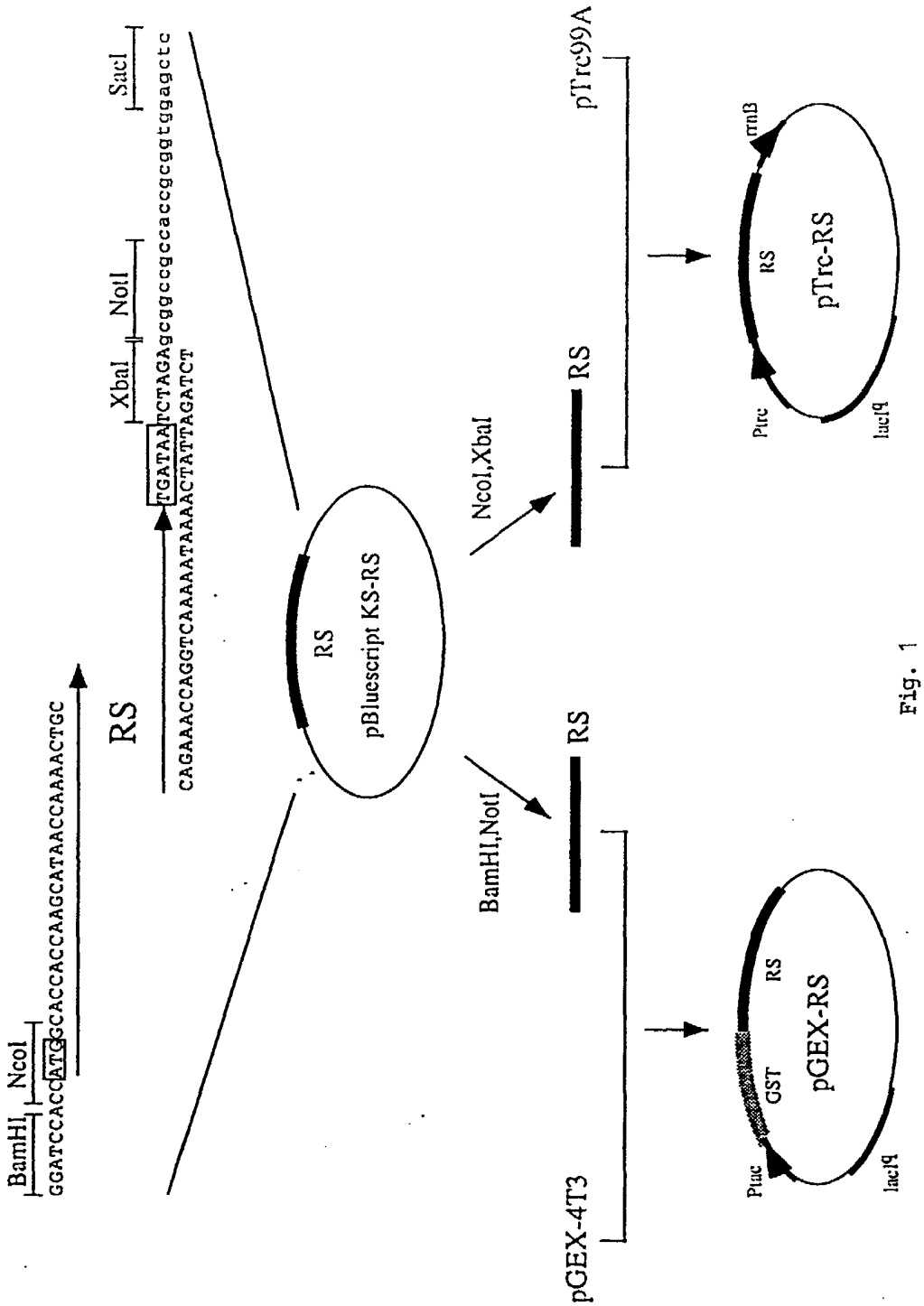
BamHI
NcoI
GGATCCACCATGGCACCACCACCAAGCATAACCAAAAACTGC
RS
CAGAAACCAGGTCAAAAATAAAACTATTAGATCT
XbaI
NotI
SacI
TGATAATCTAGAgcggccgccaccgcggtggagctc
RS
pBluescript KS-RS
NcoI,XbaI
pTrc99A
RS
pTrc-RS
RS
trnB
Ptrc
lacIq
BamHI,NotI
RS
pGEX-4T3
pGEX-RS
RS
GST
Ptac
lacIq

Fig. 1

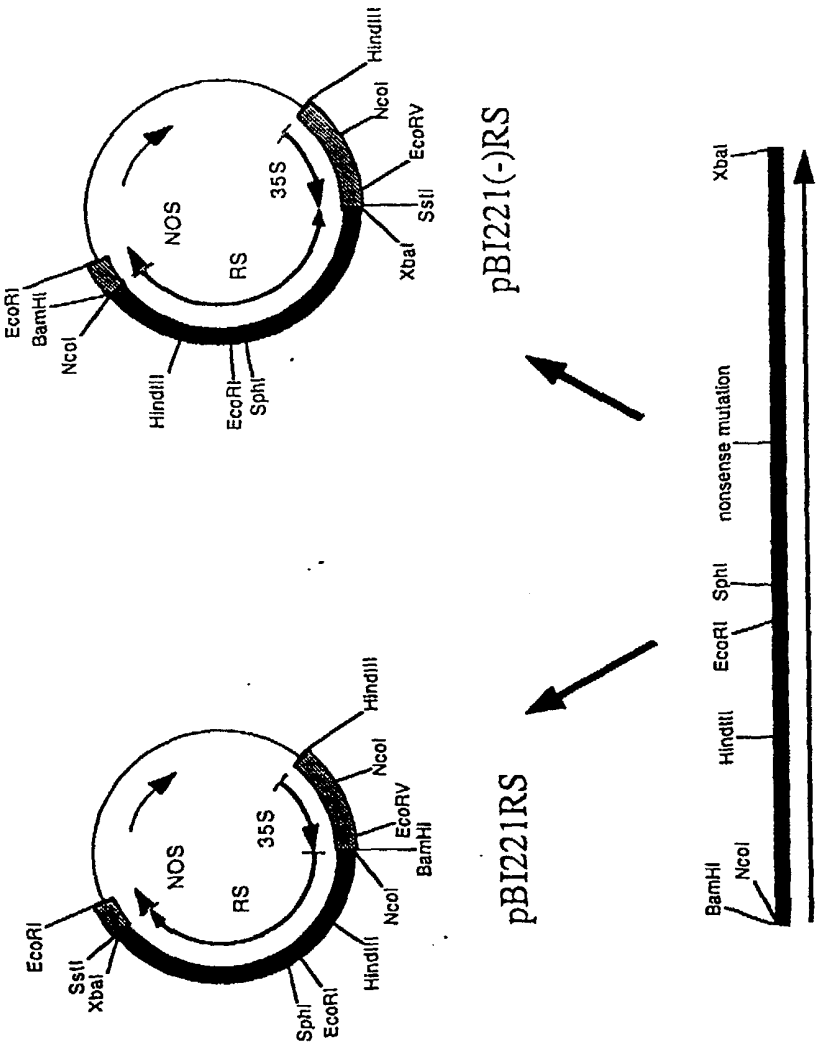
EcoRI
BamHI
NcoI
HindIII
EcoRI
SphI
NOS
RS
35S
HindIII
NcoI
EcoRV
SstI
XbaI
pBI221(-)RS
EcoRI
SstI
XbaI
SphI
EcoRI
HindIII
NcoI
BamHI
EcoRV
NcoI
HindIII
NOS
RS
35S
pBI221RS
BamHI
NcoI
HindIII
EcoRI
SphI
nonsense mutation
XbaI
RS

Fig. 2

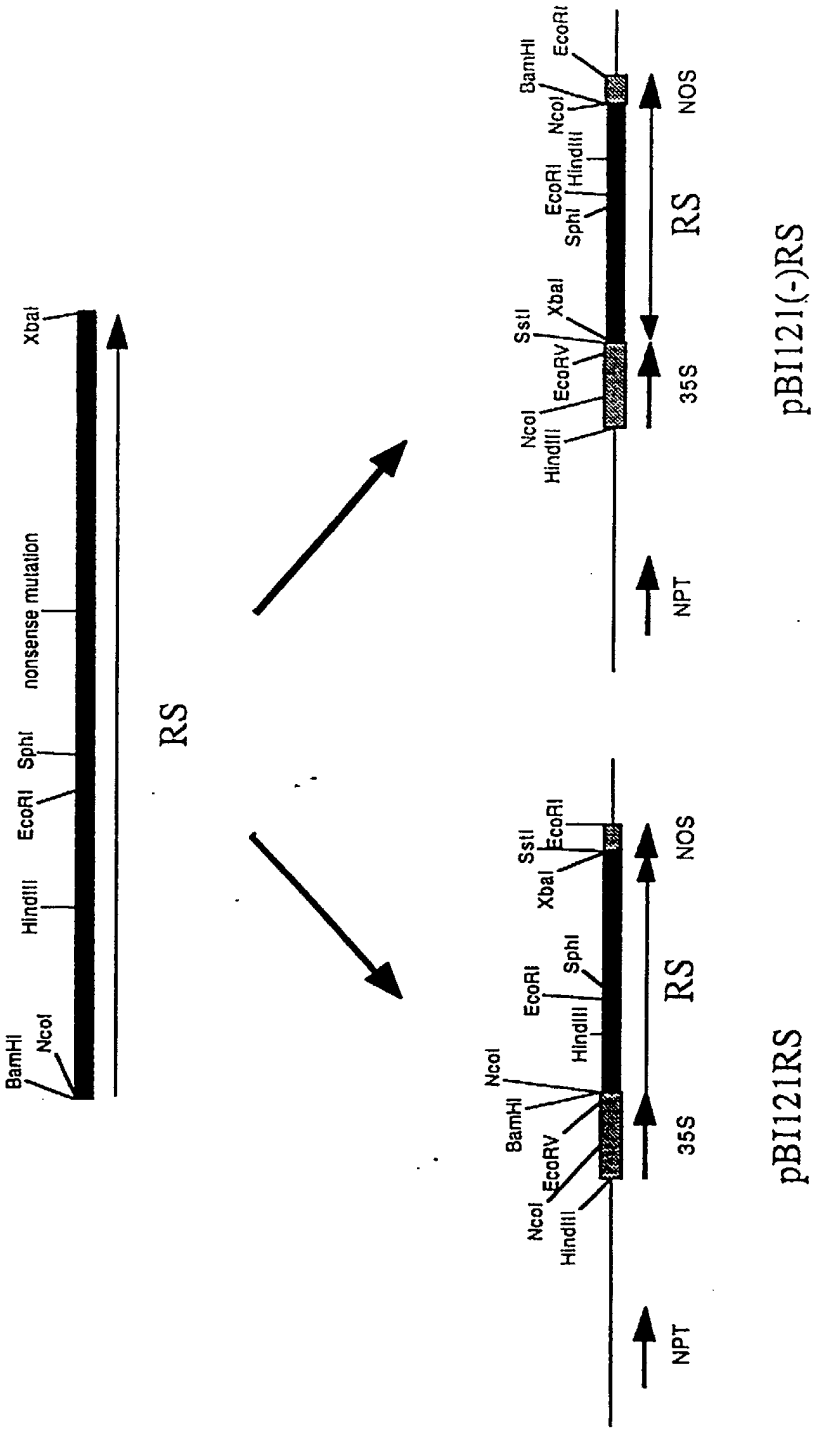
BamHI
NcoI
HindIII
EcoRI
SphI
nonsense mutation
XbaI
RS
NcoI
BamHI
EcoRV
HindIII
35S
NPT
RS
NOS
SstI
EcoRI
XbaI
SphI
EcoRI
HindIII
NcoI
BamHI
EcoRV
NcoI
HindIII
35S
NPT
RS
NOS
SstI
EcoRI
XbaI
SphI
EcoRI
HindIII
NcoI
BamHI
EcoRV
NcoI
HindIII
pBI121RS
pBI121(-)RS

Fig. 3